

MONTANA SCHOOLS OF PROMISE PERFORMANCE TRENDS AND OUTCOMES, 2009-2011



Submitted to:

Denise Juneau, Montana Superintendent of Public Instruction
Mandy Smoker Broaddus, School Improvement Grant Administrator and Program Director

Submitted by:

Christopher Lohse, Education Researcher

Spring/Summer 2013



Montana
Schools of
Promise
opi.mt.gov

Montana
Schools of
Promise

MONTANA SCHOOLS OF PROMISE PERFORMANCE TRENDS AND OUTCOMES, 2009-2011

Submitted to:

Denise Juneau, Montana Superintendent of Public Instruction
Mandy Smoker Broadus, School Improvement Grant Administrator and Program Director

Submitted by:

Christopher Lohse, Education Researcher

Spring/Summer 2013



Montana
Schools of
Promise
opi.mt.gov

TABLE OF CONTENTS

Executive Summary.....	2
Editorial Note.....	2
Introduction.....	3
Methods.....	5
Performance Rubric for the Schools of Promise.....	7
Overall Rubric Performance for the Schools of Promise.....	10
Results – Detailed.....	12
Part I. Assessment results	
Reading.....	13
Mathematics.....	19
Science.....	25
Part II. Other Academic Indicators	
Graduation rate.....	31
Attendance rates.....	33
Suspension rates.....	38
Expulsion rates.....	43
Part III. Perception Data and Findings	
Student perceptions.....	52
Teacher perceptions.....	55
Administrator and support staff perceptions.....	58
Community and parent perceptions.....	61
School board member perceptions.....	64
State team perceptions.....	66
Additional Results, Conclusions, and Further Recommendations.....	68
Appendices.....	72

EXECUTIVE SUMMARY

The Montana Office of Public Instruction (OPI), the state's executive agency for education, under the direction of the current State Superintendent of Public Instruction, Denise Juneau, decided in the spring of 2009 to apply to the United States Department of Education (USED) for a School Improvement Grant (SIG). Eligible clusters of schools in three communities eventually applied and were awarded SIG awards – Frazer, Lame Deer and Pryor, MT. Termed the Schools of Promise (SOP), the present evaluation attempts to understand whether progress has been achieved in the school improvement process undertaken by the state, districts, and participating schools. A fourteen-factor rubric was co-developed with SOP participants. The level of improvement made by each SOP site was then gauged, based on a review of available data from state-level summative assessments, locally-administered formative assessments, surveys, routinely collected data (on features like attendance, graduation rates, and discipline patterns), and approximately forty hours of stakeholder surveys, both on-site, and at intermittent state-level meetings.

Overall, the results show promising gains in Pryor (Pryor, in fact, appears to have realized the most substantive improvement in outcomes) and Frazer, with more uncertain outcomes in Lame Deer. Overall, the program appears to have offered substantial benefit. The appreciation of various stakeholders for the schools' work also appears to be increasing, while graduation rates continue to be an area of performance in need of improvement.

Each factor on the rubric (see page 10) could be gauged as green (value of 4), amber/green (value of 3), amber/red (value of 2), or red (value of 1). As a result, a maximum score of 56 (14 x 4) could be earned by a SOP participating site. Pryor garnered 75% of available points (an overall score in the green range); Frazer 73% of available points (amber/green); and Lame Deer 57% (also amber/green, but closer to amber/red). These results are summarized in the overall top-sheet for the SOP program.

The conclusions portion of the paper discusses potential next steps in appreciating the reasons for variation in the performance improvement in each community, and avenues for further inquiry.

EDITORIAL NOTE

The present paper was tasked with the work of determining, with best available evidence, whether one could credibly infer that positive, substantive changes in school performance were taking root in Montana schools that received special assistance and grant funding through a School Improvement Grant. We endeavored to gather such evidence through the end of the 2011-2012 school year. While working on the study, however, results from the 2012-2013 school year became available. The state-level summative assessments, in particular, deviate significantly from the predictions most observers would have made based on prior performance. Namely, the 2013 results were significantly lower than predicted. The present paper posits no meaningful explanation for the observed deviation from the forecasted performance.

INTRODUCTION

Montana's education system has long boasted relatively high performance on the National Assessment of Educational Progress (NAEP), the nation's only long-term reference for the academic performance of students across states and over time. Secondary graduation rates, compared to the rest of the nation, are also high for the state.

Against the backdrop of relatively high state performance, however, the state has recognized a significant challenge in the performance of American Indian students. A profound and demonstrable gap exists between the overall academic performance and graduation rate of American Indian students in Montana, and their non-Indian peers.¹

Much work has been done by the state in characterizing the dimensions of the performance challenge faced by American Indian students. In previous research done by the state, for instance, officials have demonstrated that American Indian students in the state's largest school districts tend to outperform American Indian students in reservation settings, with a notable exception in the state's most economically diverse and economically healthy reservation.²

The suppressed performance of American Indian students is not without consequences, both for the American Indian student, and for the state as a whole. When students suffer academically, they tend to drop out of school, decreasing the likelihood of career and economic success. Struggles both professionally and economically tend to have deleterious ripple effects in personal lives, as well. Thus, taken as a whole, increasing scholastic performance for the state's struggling learners is understood to be a pressing public policy dilemma with both moral and economic dimensions.

The Montana Office of Public Instruction (OPI), the state's executive agency for education, under the direction of the current State Superintendent of Public Instruction, Denise Juneau, decided in the spring of 2009 to apply to the United States Department of Education (USED) for a School Improvement Grant (SIG). The SIG funds awarded to states were part of an overall package of aid approved by the US Congress in 2009 termed the American Recovery and Reinvestment Act (ARRA).³

¹ The achievement gap on standardized assessments is reliable and consistent across a variety of psychometric instruments, most notably the National Assessment for Educational Progress (with relatively consistent effect sizes in both reading and math at age 9 and age 13), and the Montana statewide criterion referenced test. The state produces regular reports on the status of the gap. For instance, see: <http://www.opi.mt.gov/pdf/IndianEd/Data/12INEDStudentDataRpt.pdf>
<http://opi.mt.gov/pdf/IndianEd/HotTopics/10IndianEdStudentDataRpt.pdf>

² See Lohse and Ockert, September 2005. http://leg.mt.gov/content/committees/interim/2005_2006/qual_schools/staff_reports/AMERICAN_INDIAN_STUDENT_ACHIEVEMENT_IN_MONTANA_PUBLIC_SCHOOLS.pdf

³ These facts about Congressional intent become relevant later in the present report, where outcomes are matched against stated the goals of the enabling legislation.

As a pre-condition of the grant, interested school districts needed to decide upon one of four USED approved and mandated improvement models.

- **Turnaround Model:** Replace the principal, screen existing school staff, and rehire no more than half the teachers; adopt a new governance structure; and improve the school through curriculum reform, professional development, extending learning time, and other strategies.
- **Restart Model:** Convert a school or close it and re-open it as a charter school or under an education management organization.
- **School Closure:** Close the school and send the students to higher-achieving schools in the district.
- **Transformation Model:** Replace the principal and improve the school through comprehensive curriculum reform, professional development, extending learning time, and other strategies.

Under Juneau's direction, Montana OPI staff began working to identify a cohort of districts both eligible for, and interested in, the SIG resources. Ultimately, three districts were funded and enrolled in the program with USED resources. All three selected the transformation model, the most popular selection of all applicant districts in the SIG competition nationwide.

Working with representatives of the schools, districts, and communities awarded the resources, the Montana OPI collectively termed the awardee schools "Schools of Promise (SOP)."

THE SCHOOLS OF PROMISE

- *Frazer Public Schools* are located in the northeast region of the state, on the Fort Peck Reservation. The reservation is the recognized home of two American Indian tribes – the Assiniboine (or Nakoda) and the Sioux (or Dakota). It is a sparsely populated region of the state in a sparsely populated region of the nation. The closest primary trade centers – Great Falls, MT; Billings, MT; and Williston, ND – are located at distances of 305, 336, and 115 miles, respectively. The 2010 decennial census reported just 362 people in the area.
- *Pryor Public Schools* are located in south central/eastern Montana, on the Crow Reservation. The Crow Indian reservation, home to the Crow (Apsáalooke) tribe, is the largest reservation by landmass in Montana. The namesake for the community, Nathaniel Pryor, was a critical member of the Lewis and Clark expedition. The community's secondary school, Plenty Coups High School, is named for Chief Plenty Coups, Chief of the Crow during the Battle of the Little Bighorn. Pryor is located approximately 35 miles south of Billings, MT, and in the 2010 census reported a population of approximately 1,616 people.
- *Lame Deer Public Schools* are located in the southeast region of Montana, as part of the Northern Cheyenne reservation. The community of Lame Deer serves as the tribal headquarters for the reservation. The community is named for the Lakota Chief Lame Deer, killed in battle with the US army in 1877, and sits approximately 100 miles east of Billings, MT. The 2010 decennial census reported 2,052 people living in the region.

The present study attempts to discern whether interventions and changes made as a result of SOP participation have culminated in improvements to education practice and policy within the participating communities and schools, and more importantly, whether those improvements have resulted in better outcomes – both academic and non-academic – for learners.

METHODS

Untangling the question of whether investments and interventions have yielded educational benefits is not as simple as comparing test scores before the SOP program started with today's results. Consider the following situations which made this difficult.

- It could be that test scores increased over the time interval, but it's possible that scores were improving at a faster rate prior to participation in the SOP cohort.
- It is possible that declining scores could be seen as a success, if SOP interventions appear to slow the decline. The only way of knowing whether decline was likely is to look at statistically alike comparison groups.
- Scores can rise, often leading people to conclude that improvements in a given area are working. But if the local increase is a much smaller increase than the rest of the state, doubts are raised about the quality of the interventions.

To avoid such common misinterpretations of education outcomes, a significant amount of historic data, and data from other Montana school sites – quasi-experimental, “control” sites, were analyzed in the present evaluation.

Moreover, academic outcomes can serve as lagging indicators of underlying performance. In other words, an improving school may not see immediate gains in estimates of student scholastic performance. To appreciate whether the general climate and atmosphere in the school appears poised to sustain greater scholastic excellence, a number of interviews were conducted with diverse stakeholders and analyzed much more than student assessment scores.

Even with such analytical safeguards in place, it is always difficult to attribute any observed improvement in a complex system to a discreet set of underlying interventions. The present report does not attempt to draw scientific, causal relationships between SOP interventions and measured outcomes. Instead, the purpose is attempt to draw credible inferences about the success of the program from the review of the data – statements that appear logical from an unbiased and thorough review of available information. In essence, this is an evaluative study that attempted to answer whether there are credible signs of progress in the school sites, rather than a research study, that would attempt to answer how and why certain practices were gaining foothold and demonstrating success in some school sites, and not in others. In the discussion of results, and the attempt at some conclusions, however, some potential hypotheses are offered as to why some features may be occurring, and suggestions are made regarding

research studies that may be eventually pursued by interested parties to validate or invalidate those hypotheses.

Additionally, a goal of the Schools of Promise program itself was to develop the capacity of the local districts to monitor performance and course correct as necessary to drive toward higher levels of performance. Evaluation should not be a process that necessarily requires external or specialized expertise, but rather a process that should feel approachable and understandable by a wide variety of school stakeholders. Thus, there was an attempt to build a simple-to-use, easily replicated rubric and accompanying dashboard for understanding school performance.

Many of the performance features incorporated into the rubric and dashboard are easily recognized, common descriptors of school performance: summative, standardized test scores; graduation and attendance rates; and rates of corrective disciplinary action. However, because of conversations with various school teams about other features that they felt strongly “mattered” in schools, data about how engaged students were in school, and how various stakeholders perceived the school was also incorporated. To examine these less familiar constructs, survey data, extensive interviews, and techniques drawn from or inspired by the science of complex-adaptive systems were used.

The resulting rubric is attached in Figure 1. It should be noted that the rubric is for a set of “outcomes” about the school, and is thus limited in its ability to guide improvement as a stand-alone document. Discussions with the Schools of Promise teams reinforced how measures of 1) school inputs, 2) school processes and 3) school outcomes were all relevant, and that this was an attempt to generate a look at the third set of school outcomes measures. The Schools of Promise were already using a number of input and process measures – including, notably, a set of input and process measures inspired by Larry Lazotte’s work citing correlates of effective schools that has been modified to incorporate ways of examining performance in American Indian contexts and communities – and there was not an intention to introduce another, potentially confusing analytic framework for considering inputs and processes.

Figure 1. Performance Rubric for the Schools of Promise

(also included in Appendix A)

	Red	Amber/Red	Amber/Green	Green
Reading improvement	The percentage of students who are scoring proficient or higher has declined. The percentage of students performing at the novice level has grown.	Overall, there may be little to no improvement in the percentage of students scoring proficient or higher. There may, however, be a decline in the percentage of students scoring at the novice level.	More students are scoring proficient, and fewer students are scoring at the novice level. Improvement, however, may not outpace comparison schools.	More students are scoring proficient, and fewer students are scoring at the novice level, and the improvement exceeds the rate of improvement in comparison schools.
Mathematics improvement	The percentage of students who are scoring proficient or higher has declined. The percentage of students performing at the novice level has grown.	Overall, there may be little to no improvement in the percentage of students scoring proficient or higher. There may, however, be a decline in the percentage of students scoring at the novice level.	More students are scoring proficient, and fewer students are scoring at the novice level. Improvement, however, may not outpace comparison schools.	More students are scoring proficient, and fewer students are scoring at the novice level, and the improvement exceeds the rate of improvement in comparison schools.
Science improvement	The percentage of students who are scoring proficient or higher has declined. The percentage of students performing at the novice level has grown.	Overall, there may be little to no improvement in the percentage of students scoring proficient or higher. There may, however, be a decline in the percentage of students scoring at the novice level.	More students are scoring proficient, and fewer students are scoring at the novice level. Improvement, however, may not outpace comparison schools.	More students are scoring proficient, and fewer students are scoring at the novice level, and the improvement exceeds the rate of improvement in comparison schools.
Graduation rate improvement	The percentage of students who are graduating has declined.	There is no improvement in the percentage of students graduating, or irregular and erratic patterns of improvement over the time interval considered.	There is modest improvement in the percentage of students graduating, or generalized trend improvement over the time interval examined.	Higher percentages of students are graduating from high school.
Attendance improvement	Demonstrably fewer students are attending school, and the rate of attendance decline is steeper than in comparison groups.	There is no improvement in the percentage of students graduating, or irregular and erratic patterns of improvement over the time interval considered.	Attendance trends have improved, but as not as sharply as in comparison schools (eg state and/or historically and demographically similar schools)	Attendance trends have improved, and at rates that exceed the attendance improvement in comparison schools.
Suspension improvement	More students are being suspended, and at a rate that exceeds comparison schools.	Suspension rates are increasing, but not at a faster rate than comparison groups.	Suspension rates are declining, but not at a rate that outpaces comparison groups.	Suspension rates are declining at a rate that exceeds the rate of decline in comparison schools.
Expulsion improvement	More students are being expelled, and the rate is increasing faster than comparison groups.	Expulsion rates have not improved, and/or are erratic over the time interval. Expulsion rates may be higher, but may not be increasing at a rate that exceeds that of comparison schools.	Expulsion rates are declining, generally, but not in a way that clearly outpaces comparison schools.	Expulsion rates are declining at a rate that exceeds the rate of decline in comparison schools.

Figure 1 (cont.) Performance Rubric for the Schools of Promise
(also included in Appendix A)

	Red	Amber/Red	Amber/Green	Green
Student engagement improvement	Scores of student engagement, as measured and visualized, are declining.	Scores of student engagement are not improving, or are declining, but not as rapidly as comparison schools.	Student engagement rates are increasing, but at a rate that is slower than comparison groups.	Student engagement rates are increasing at a rate that exceeds the increases seen in comparison groups.
	Red	Amber/Red	Amber/Green	Green
Perception improvement (students)	The stakeholder group perceives that the school is in decline, performing worse than it was historically.	The stakeholder group perceives little improvement in the school, and is not optimistic about the prospects for future success.	The stakeholder group may or may not perceive improvement in the school, but is optimistic about the prospects for future success.	The stakeholder group perceives improvement in the school, and is optimistic about the prospects for future success.
	Red	Amber/Red	Amber/Green	Green
Perception improvement (teachers)	The stakeholder group perceives that the school is in decline, performing worse than it was historically.	The stakeholder group perceives little improvement in the school, and is not optimistic about the prospects for future success.	The stakeholder group may or may not perceive improvement in the school, but is optimistic about the prospects for future success.	The stakeholder group perceives improvement in the school, and is optimistic about the prospects for future success.
	Red	Amber/Red	Amber/Green	Green
Perception improvement (administrators and support staff)	The stakeholder group perceives that the school is in decline, performing worse than it was historically.	The stakeholder group perceives little improvement in the school, and is not optimistic about the prospects for future success.	The stakeholder group may or may not perceive improvement in the school, but is optimistic about the prospects for future success.	The stakeholder group perceives improvement in the school, and is optimistic about the prospects for future success.
	Red	Amber/Red	Amber/Green	Green
Perception improvement (community)	The stakeholder group perceives that the school is in decline, performing worse than it was historically.	The stakeholder group perceives little improvement in the school, and is not optimistic about the prospects for future success.	The stakeholder group may or may not perceive improvement in the school, but is optimistic about the prospects for future success.	The stakeholder group perceives improvement in the school, and is optimistic about the prospects for future success.
	Red	Amber/Red	Amber/Green	Green
Perception improvement (board members)	The stakeholder group perceives that the school is in decline, performing worse than it was historically.	The stakeholder group perceives little improvement in the school, and is not optimistic about the prospects for future success.	The stakeholder group may or may not perceive improvement in the school, but is optimistic about the prospects for future success.	The stakeholder group perceives improvement in the school, and is optimistic about the prospects for future success.
	Red	Amber/Red	Amber/Green	Green
Perception improvement (state)	The stakeholder group perceives that the school is in decline, performing worse than it was historically.	The stakeholder group perceives little improvement in the school, and is not optimistic about the prospects for future success.	The stakeholder group may or may not perceive improvement in the school, but is optimistic about the prospects for future success.	The stakeholder group perceives improvement in the school, and is optimistic about the prospects for future success.

In reading the rubric performance standards for each criterion, one will notice the requirement of referencing performance against relevant comparison groups. As mentioned earlier in this report, in order to increase the credibility of the inferences in the present evaluation, the identification of quasi-experimental “controls” was necessary. Fortunately, two reasonable comparison groups avail themselves for the purpose of this evaluation.

First, all Schools of Promise had previously been engaged in a state-led and orchestrated continuous improvement process. Termed “A Call to Greatness,” all schools that had been labeled as “persistently in need of improvement” were convened by the state, provided additional school improvement professional development, and coached. Other than the Schools of Promise, all other Call to Greatness participants were also located on American Indian reservations within the state. In fact, every state reservation, except the Flathead Reservation (which has comparatively high levels of American Indian student outcomes), sent delegates to the Call to Greatness conventions. Because the other Call to Greatness (CTG) schools were so racially and economically alike; because they, too, had been engaged in a state-led improvement process; and because several – but not all – of the theories of action undergirding their improvement processes were alike, they made a suitable reference group. It is important to note, however, that as a pre-condition for grant eligibility, all of the schools in the SIG program had to be the very lowest performing in the state. The Schools of Promise, then, were not randomly selected from the Call to Greatness family of schools – they were selected precisely because they struggled even more than other struggling schools. It was thus expected that their performance would lag behind the CTG schools; the interest was in whether the rate of improvement in SOP schools exceeded the rates of improvement in CTG schools.

Second, there were two other school sites that were eligible for SIG support – Hays-Lodge Pole Schools, and Lodge Grass Schools. In 2009, both pursued grant participation for a period of time before ultimately withdrawing as grant participants. Nevertheless, their kinship in meeting grant eligibility requirements made them a useful comparison group, as well.

Because the data collections for the Schools of Promise vary somewhat from overall school data collections in the state, it was not possible to construct a reference or comparison group for all strands of the rubric. In these instances, the control became the state as a whole, or references to comparison groups were extracted from the rubric.

Results – Overall

Figure 2. Overall Rubric Performance for the Schools of Promise

	SCHOOLS of PROMISE	FRAZER	LAME DEER	PRYOR
Reading improvement				
Mathematics improvement				
Science improvement				
Graduation rate improvement				
Attendance improvement				
Suspension improvement				
Expulsion improvement				
Student engagement improvement				
Perception improvement (students)				
Perception improvement (teachers)				
Perception improvement (administrators and support staff)				
Perception improvement (community)				
Perception improvement (board members)				
Perception improvement (state)				

Please note: Student engagement improvement was measured and visualized using a construct that has not yet been fully approved. The rating is indicated here, but engagement measures are excluded from the discussion.

Figure 2 represents the rubric scores for each criterion for four different units of analysis. The first column shows the overall ranking for the SOP program. The second column shows the ranking for Frazer Schools. The third column shows the ranking for Lane Deer Schools. The fourth column shows Pryor Schools' ranking.

Overall, the Schools of Promise program appears to be adding value and yielding improvement. To generate a quantitative value for the SOP overall, and for each school, each of the rubric values was coded to a corresponding numeric value (red = 1 point; amber/red = 2 points; amber/green = 3 points; and green = 4 points). A maximum score of 164 points would be possible, if all values for all sites had been green -- 14 rubric rows X 12 points per row (4 for each green score at each school site, with one missing value for one school, worth 4 points). The actual points earned by the various school sites were summed and divided by the 164 point total.

The straightforward calculation for the Schools of Promise as a whole was thus





$$\frac{\text{Points earned (Schools of Promise)}}{\text{Points possible (Schools of Promise)}} = \frac{115}{164} = 70.1\%$$

Values between 75% and 100% were taken to be green; values between 50% and 74.9% were taken to be amber-green; values between 25% and 49.9% were taken to be amber-red; and values below 25% were taken to be red.

For the school sites themselves, the calculation was precisely the same, but points earned and points possible were for those sites, exclusively. The results appear in Figure 3 below.

From the school level reviews, it is evident that improvement in the SOP program is not evenly distributed. Based on evidence reviewed in the evaluation process, Pryor and Frazer appear to demonstrate improvement on a variety of indicators, while improvement in Lane Deer is less certain, and more imperiled. In the conclusions section of the paper, there are hypotheses for the observed variation in school-level outcomes and suggestions regarding possible research studies to elucidate support (or invalidation) of the initial hypotheses.

Figure 3. Overall outcomes in the Schools of Promise

Schools of Promise	70.1%	
Frazer	73.2%	
Lane Deer	57.1%	
Pryor	80.8%	

RESULTS – DETAILED

PART I. STUDENT ASSESSMENT OUTCOMES

Two psychometric instruments were primarily used to gauge and determine the level of performance in each of the assessment domains: the state-level criterion referenced test (CRT), termed the Montana Comprehensive Assessment System (MontCAS); and the Northwest Evaluation Assessment (NWEA) benchmark assessments, termed the Measures of Academic Progress (MAP) tests.

Discerning readers will note, however, that the terminology of the rubric is only appropriate when examining MontCAS scores. MAP data was used to inform the conversations about assessment, but not the overall performance rank for the school. The reason for MAP exclusion from the overall performance rank were manifold. For both Frazer and Pryor, the deployment of the MAP assessment was new in 2011, and teachers and students were still learning how best to use the results to inform instruction. Moreover, the use of the assessments was still uneven in the school sites. Ideally deployed, there would be an assessment score for each child in the fall, winter, and spring. In fact, the number of students for whom all three scores are available were precious few (and sample sizes by grade were already small in the Schools of Promise). Additionally, not every site administered the test in each recommended trimester, making meaningful roll-ups for the benchmark NWEA assessment outcomes difficult.

Nevertheless, reactions to the NWEA assessment were generally positive, with teachers pleased to have more ongoing information about their student's academic progress in reading and mathematics. Moreover, to the extent that statements can be made about changes in MAP outcomes, it is clear that more students in the SOP overall are reaching the academic standards typically associated with their grade-level and chronological age.

In the future, modifying the rubric to reflect a review of both the interim and summative assessment outcomes is suggested. For now, the basis of the assessment statements is based, unfortunately, exclusively on the performance on the MontCAS assessment.

READING IMPROVEMENT

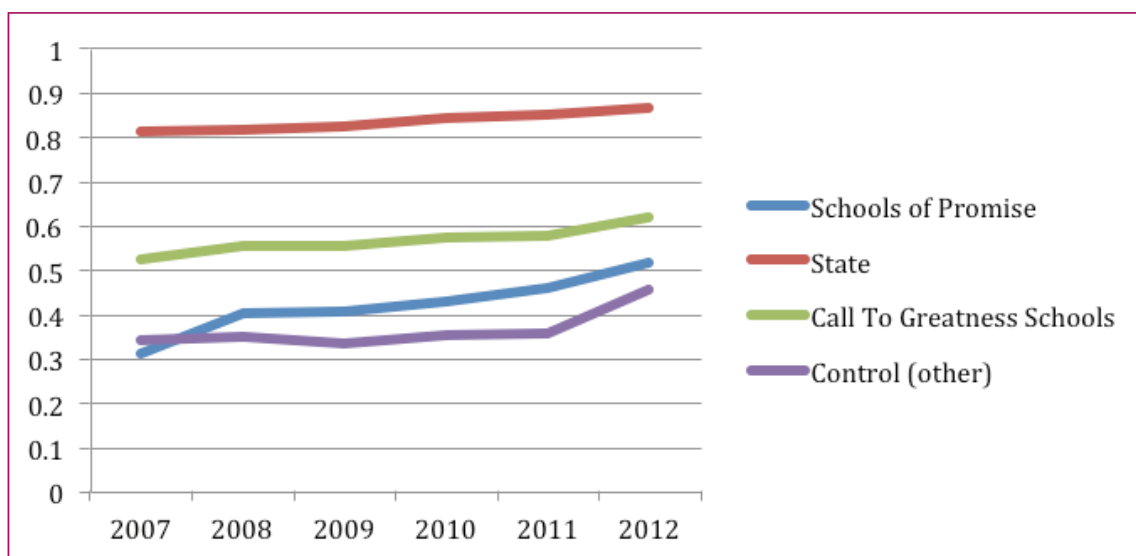
Figure 4. Reading Assessment Improvement in the Montana Schools of Promise

Overall	Frazer
	Lame Deer
	Pryor

As shown in Figure 4, all sites met the criteria for a “green” or “amber green” rubric rating. Improvement in reading between 2009 and 2012 proceeded at a rate that was faster than the state, and faster than at least one of the comparison groups. A deeper interrogation of the available data justifies these high ratings.

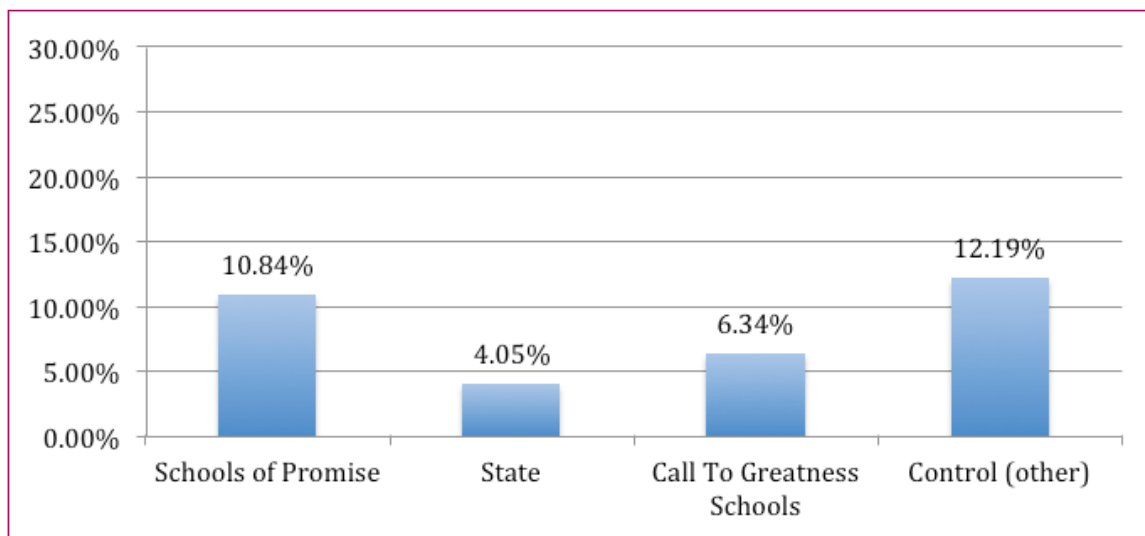
The reading assessment statements are based on the performance of students in grades 3-8 and grade 10 on the MontCAS assessment.

Figure 5. Percent of students achieving proficient or advanced levels of performance on Reading assessments of the MontCAS, 2007-2012



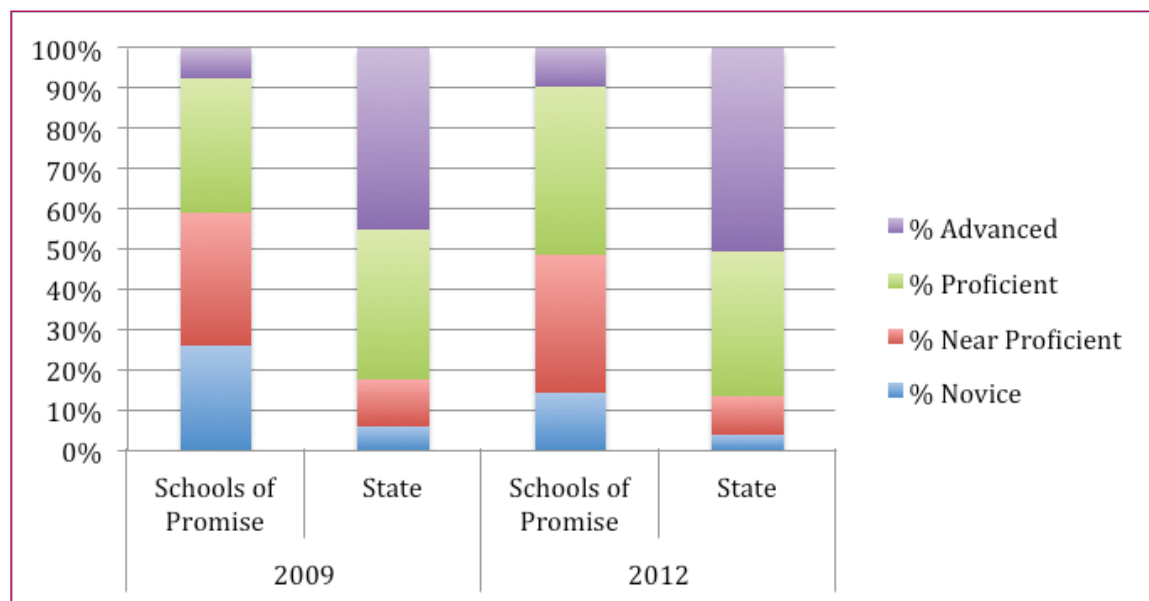
The Schools of Promise have demonstrated steeper rates of growth in reading over the time interval from 2009 to 2012 than either the Call to Greatness schools or schools in the state as a whole (Figure 5). Because of a sharp increase in performance in the Lodge Grass and Lodge Pole schools, Schools of Promise did not exceed their performance growth, but much of the growth by the Lodge Grass and Lodge Pole comparison schools occurred in the early grades only and were not SIG eligible schools, and was concentrated in the last academic year. Moreover, there had been significant academic growth in the Schools of Promise that preceded the SIG interventions. But overall, these performance features align with a “green” rating, as growth was significant and outsized when compared with two reference groups over the grant interval, made clearer in Figure 6.

Figure 6. Reading growth by school groupings, 2009-2012



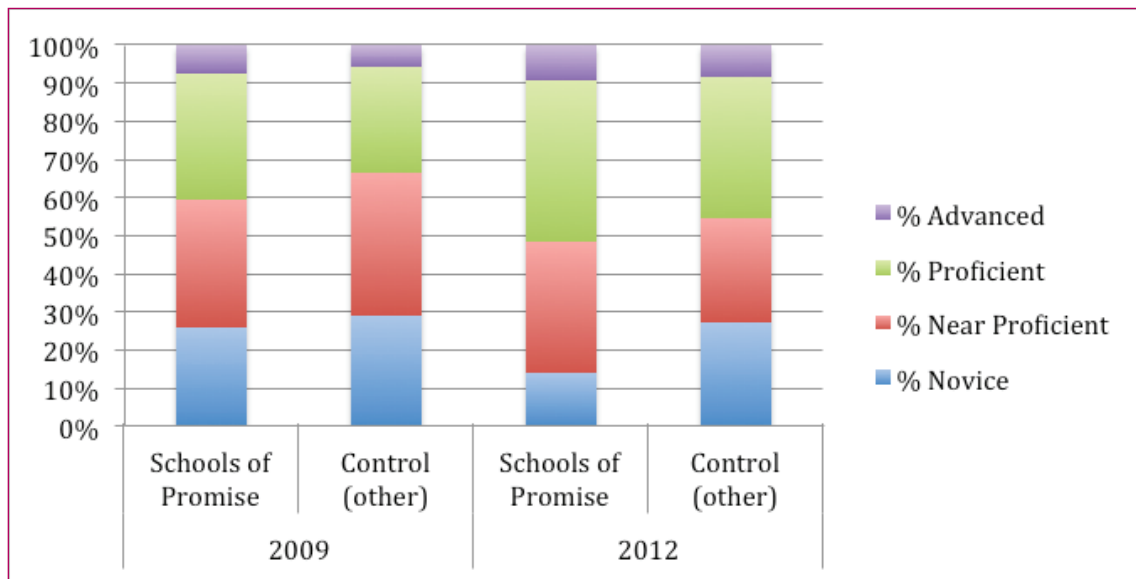
In Figure 6, the performance trend is clearer. Significant growth, outpacing both the state and the Call to Greatness Schools, requires a green determination.

Figure 7. Performance distribution on the MontCAS Reading Assessment, 2009 and 2012



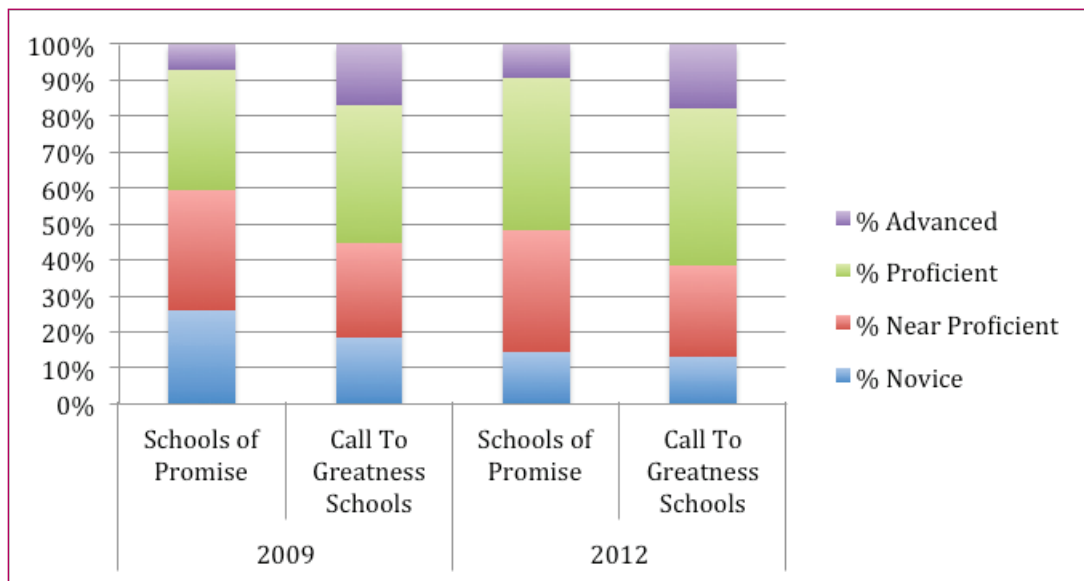
In addition to the growth that occurred in the percentage of students achieving proficiency or higher on the MontCAS, Figure 7 shows a very significant decline in the percentage of “novice” students. Indeed, the percentage of students scoring at the novice level was nearly cut in half over the time interval of the grant. The strong growth in the percentage of students “nearing proficiency” in SOP must be considered a grant success.

Figure 8. Performance distributions on the MontCAS Reading Assessment, 2009 and 2012



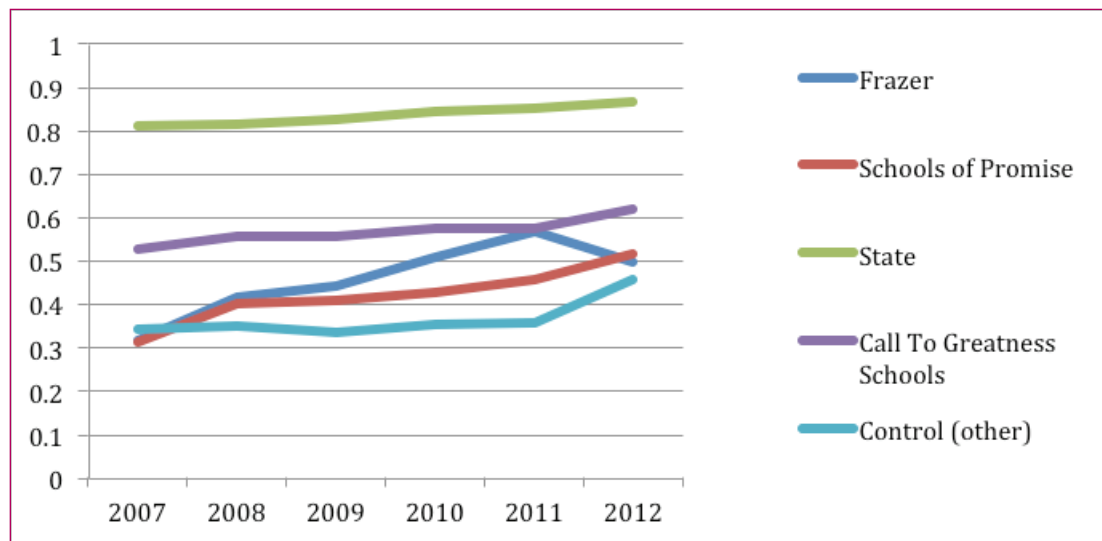
The Schools of Promise fare equally well in distribution analyses when examined against other SIG-eligible schools (see figure 8). The decline in the percentage of novice students in Lodge Grass and Lodge Pole schools was not nearly as robust as it was in the Schools of Promise.

Figure 9. Performance distributions on the MontCAS Reading Assessment, 2009 and 2012



Call to Greatness Schools generally shrank their novice population, but again failed to keep up with the rate of novice decline in the Schools of Promise (see Figure 9). This performance feature again bolsters support for a green rating.

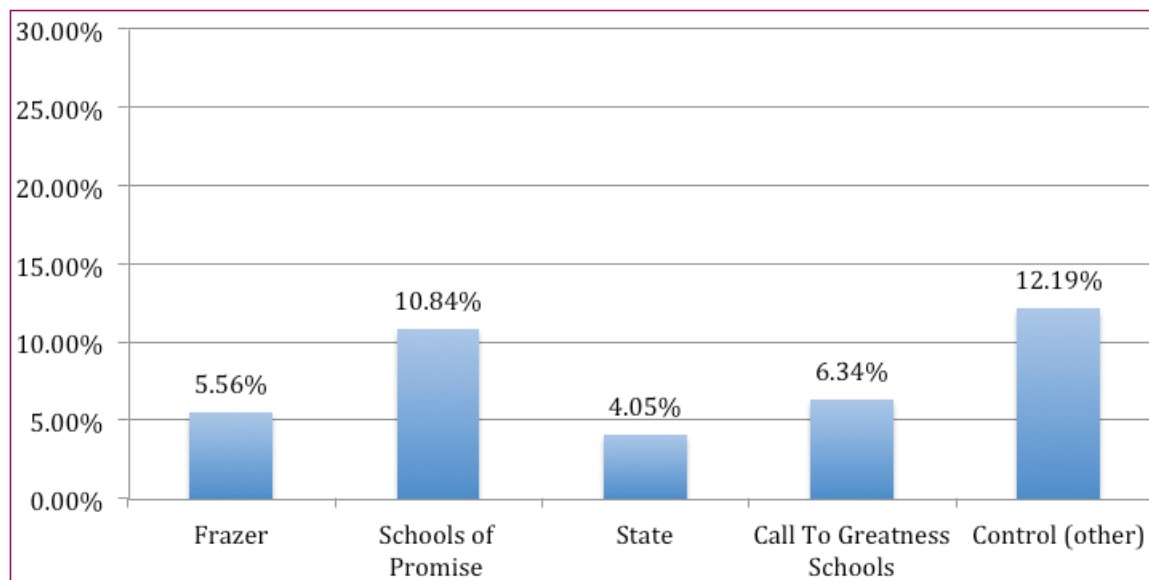
Figure 10. Percent of students achieving proficient or advanced levels of performance on Reading assessments of the MontCAS, 2007-2012



Frazer demonstrated strong reading growth until the most recent MontCAS administration (Figure 10).

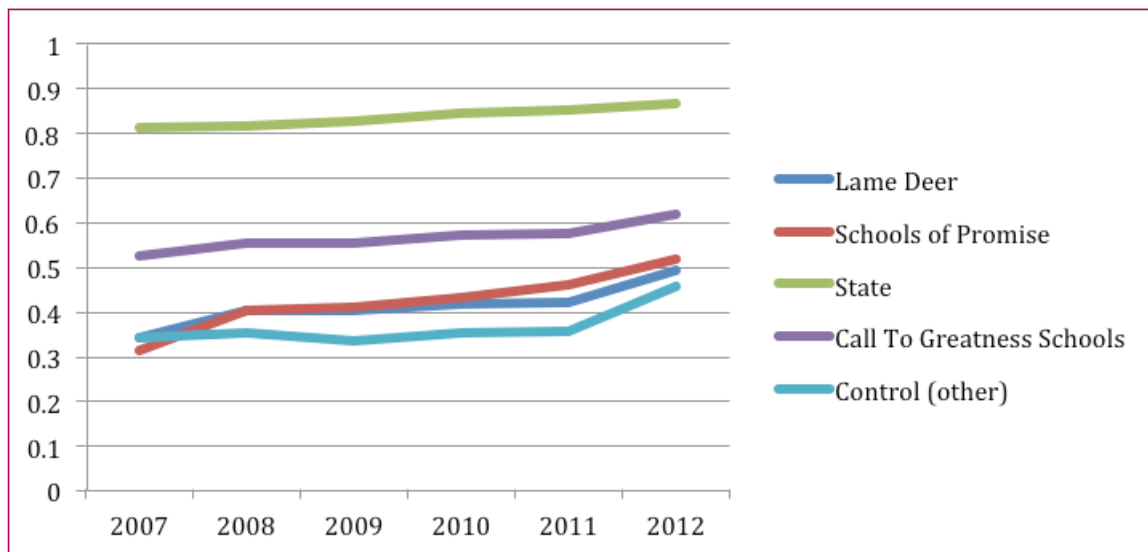
Despite their one-year decline, however, growth for the overall grant period remained higher than the state, and not statistically different from the Call to Greatness Schools, as indicated by Figure 11.

Figure 11. Reading growth by school groupings, 2009-2012



As a result of the one-year decline, however, an amber-green determination was warranted for Frazer schools.

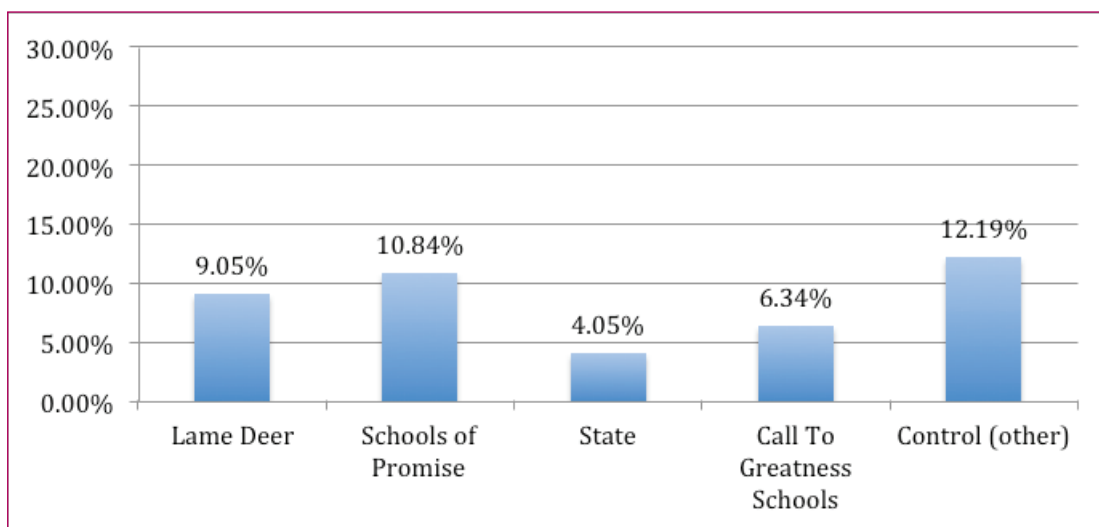
Figure 12. Percent of students achieving proficient or advanced levels of performance on Reading assessments of the MontCAS, 2007-2012



Lame Deer Reading growth was consistent over the time interval as shown in Figure 12.

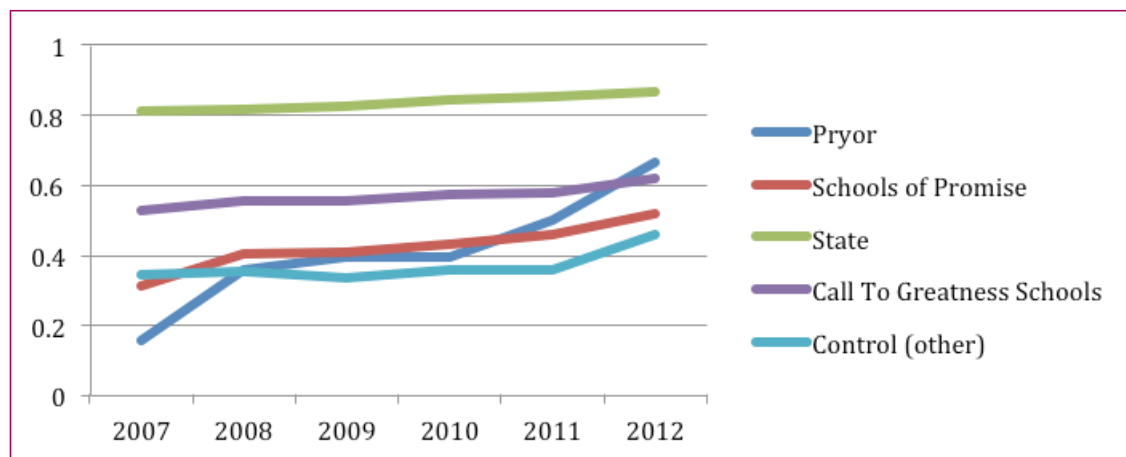
Indeed, as is evident in Figure 13, the reading growth during the grant administration exceeded both Call to Greatness schools, and the state as a whole.

Figure 13. Reading growth by school groupings, 2009-2012



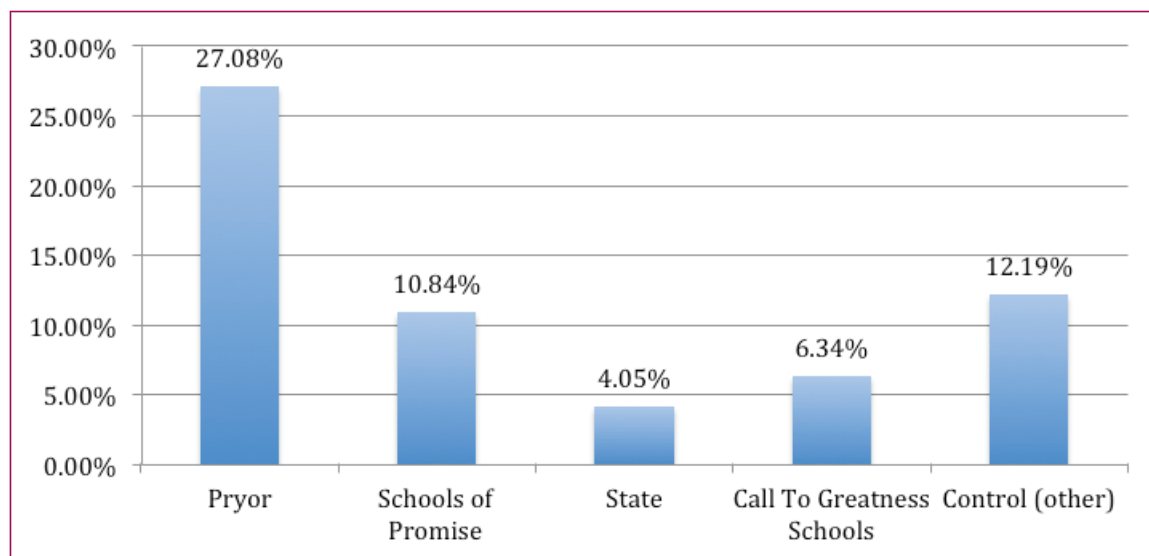
Because of their comparative success at increasing academic outcomes in reading, Lame Deer garnered a green rating.

Figure 14. Percent of students achieving proficient or advanced levels of performance on Reading assessments of the MontCAS, 2007-2012



Pryor demonstrated the most significant reading gains of any of the Schools of Promise. In fact, in the most recent MontCAS administration, Pryor actually exceeded the overall performance of the other Call to Greatness schools (see Figure 14). It is worth reiterating that Pryor was selected for SIG participation precisely because it was one of the lowest performing in the Call to Greatness cohort, so exceeding the group's overall performance represents sharp growth. It is also noteworthy to acknowledge that Pryor was the only K-12 system identified in the SIG cohort (Frazer was 9-12 and Lame Deer was 7-12).

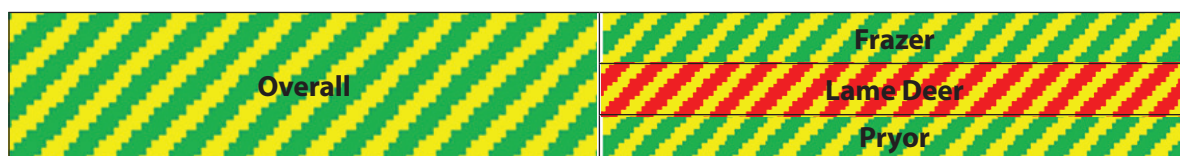
Figure 15. Reading growth by school groupings, 2009-2012



The sharp growth of Pryor over the life span of the grant is made even clearer in Figure 15. Far outpacing all comparison groups, a rating of green was warranted for Pryor Schools.

MATHEMATICS IMPROVEMENT

Figure 16. Mathematics Assessment Improvement in the Montana Schools of Promise

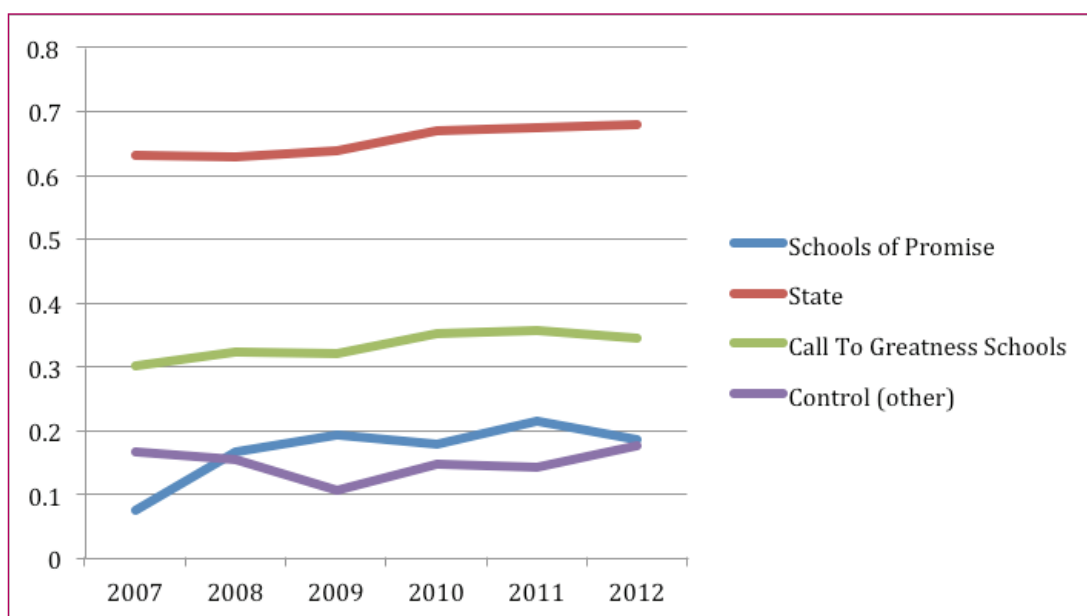


Trends in mathematics performance were less overwhelmingly positive than the trends for reading. In fact, when averaged as a large group, performance actually declined in mathematics over the interval from 2009-2012. Contributing factors may include the fact that greater resources were employed towards literacy and English Language Arts in year one of the SIG grant as part of the overall SOP strategy and three year plan as well as the fact that more significant and abrupt changes were implemented in the sequence of the 7-12 mathematics program.

The mathematics performance patterns were erratic, not in linear decline. Moreover, like in reading, the percentage of students performing at the novice level decreased significantly, and that fact, coupled with modest growth in Frazer and relatively significant growth in Pryor, yielded an overall amber green rating.

The mathematics assessment statements are based on the performance of students in grades 3-8 and grade 10 on the MontCAS assessment.

Figure 17. Percent of students achieving proficient or advanced levels of performance on Mathematics assessments of the MontCAS, 2007-2012



From Figure 17 (above) and Figure 18 alone, it seems that SIG interventions may have slowed earlier mathematics growth in the SOP, and that the SOP are outpaced by growth in at least one of the controls

and the overall state. But a deeper review of both the performance distributions and the site-level performance counters such a bleak view of the success of math interventions.

Figure 18. Mathematics growth by school groupings, 2009-2012

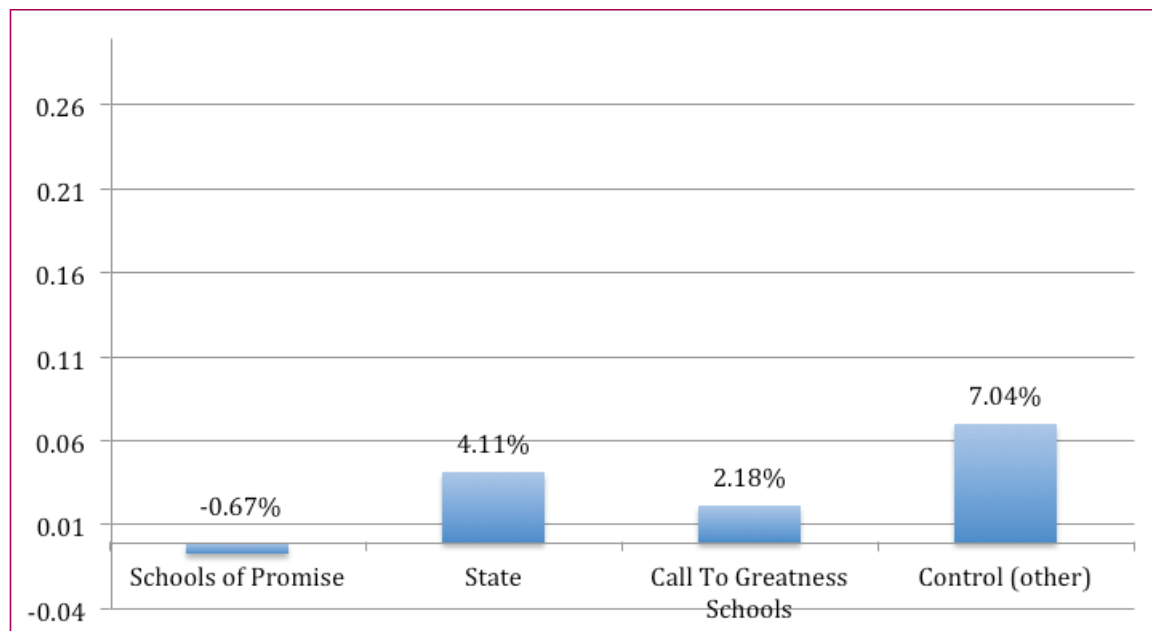
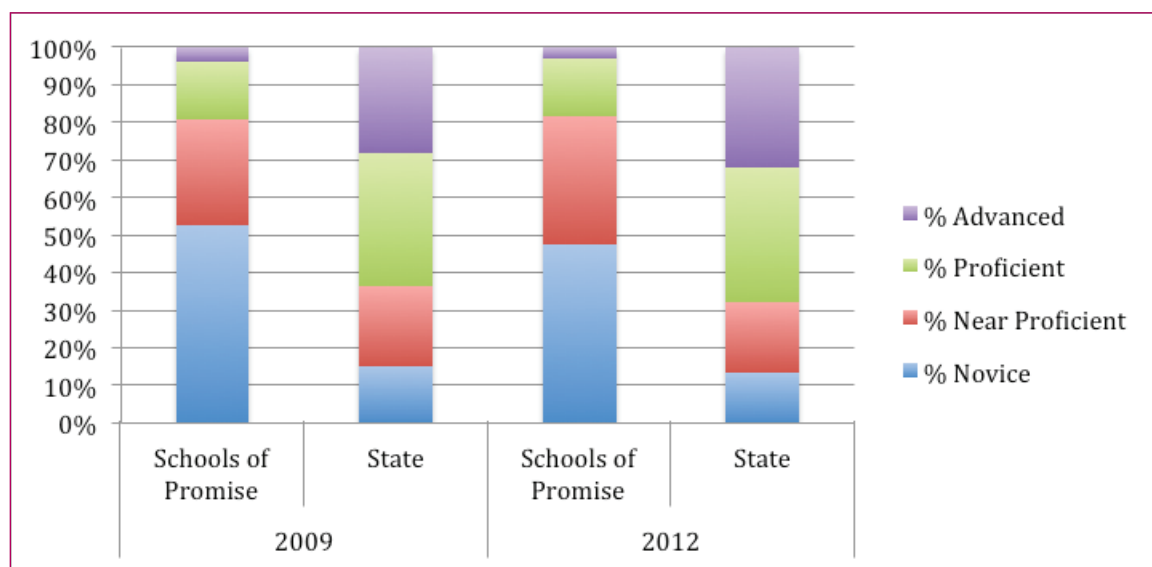
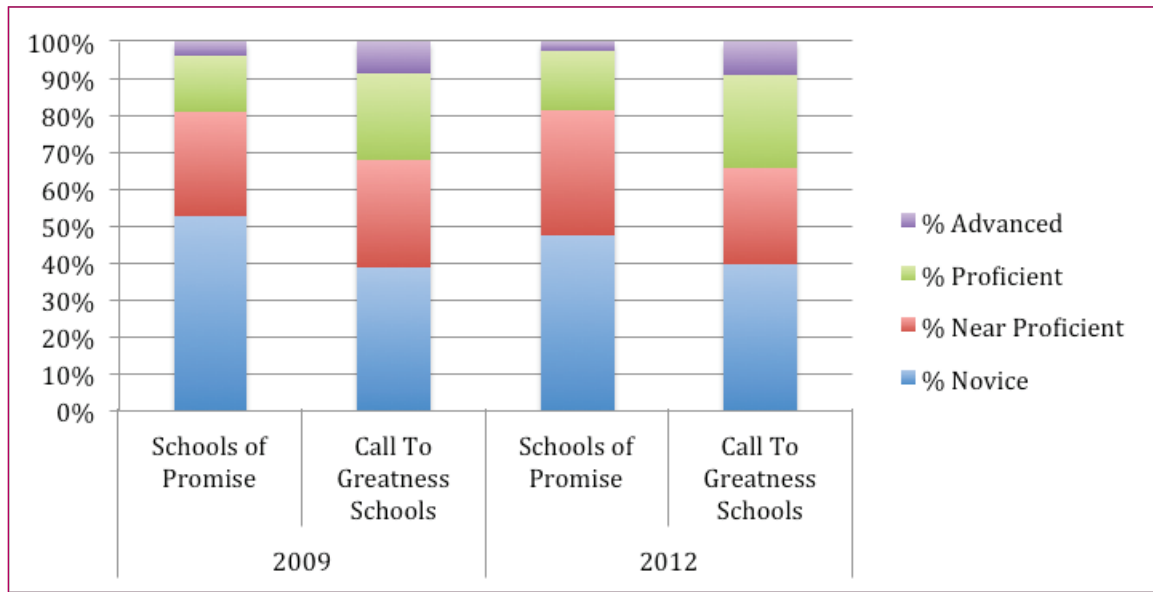


Figure 19. Performance distributions on the MontCAS Mathematics Assessment, 2009 and 2012



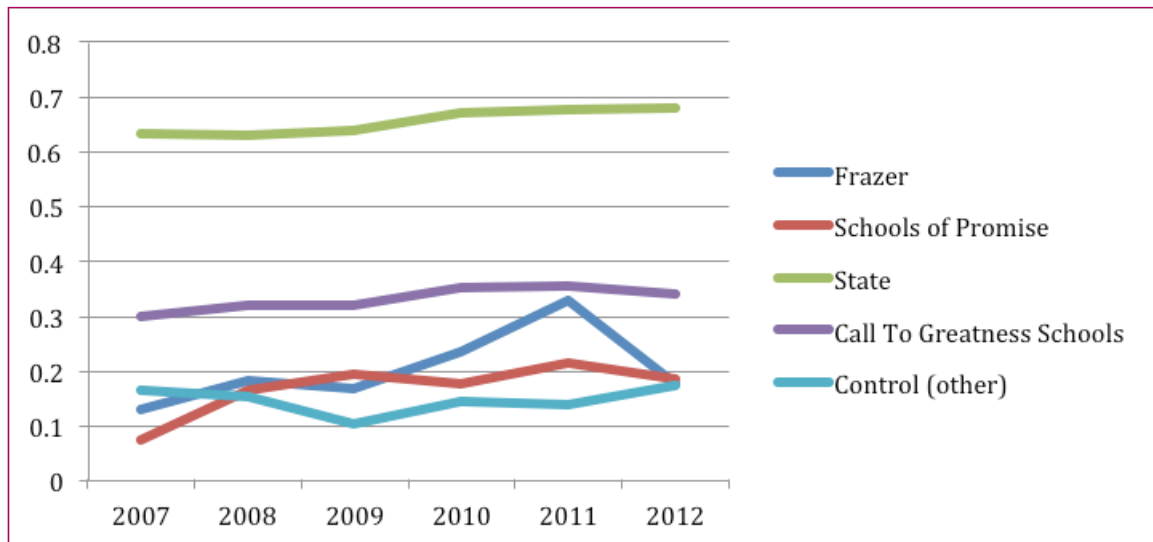
The percentage of students performing at the novice level shrunk significantly, going from about 52% in 2009 to about 47% in 2012 in SOP (see Figure 19). The achievement far outpaced the same measure for the state as a whole. This positive indicator factored in the amber-green rating for overall mathematics performance.

Figure 20. Performance distributions on the MontCAS Mathematics Assessment, 2009 and 2012



The positive changes in the performance distributions are even more significant when compared with other Call to Greatness Schools (Figure 20). The CTG control schools did not decrease their percentage of novice performers at all between 2009 and 2012.

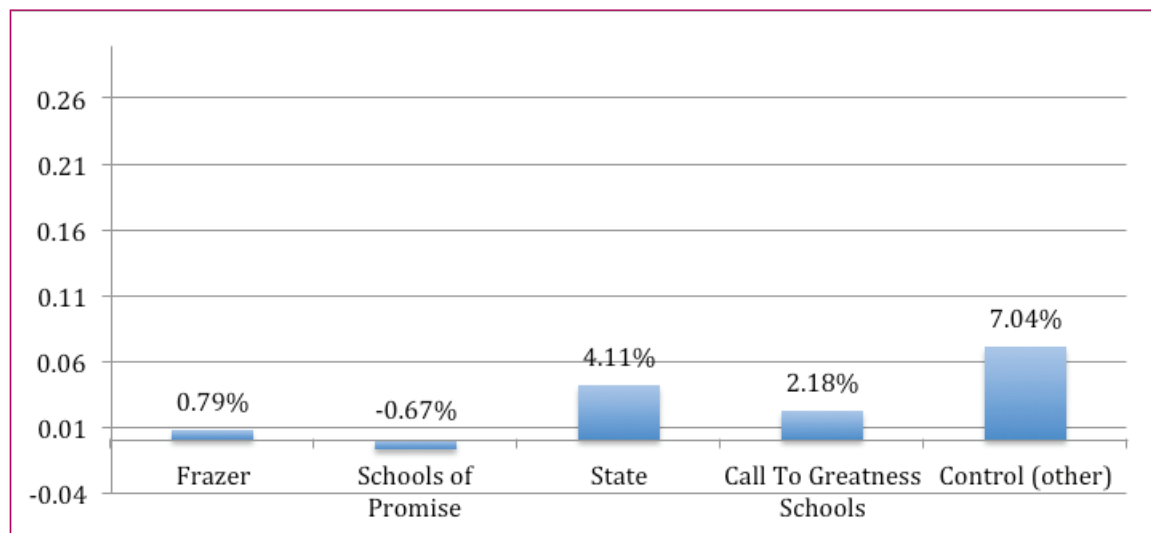
Figure 21. Percent of students achieving proficient or advanced levels of performance on Mathematics assessments of the MontCAS, 2007-2012



Looking at the annual performance for Frazer in Figure 21, one again sees how complicated mathematics performance looks for the SOP. A precipitous drop in the most recent test administration followed very strong, two-year mathematics growth at the start of the grant. School officials are hopeful

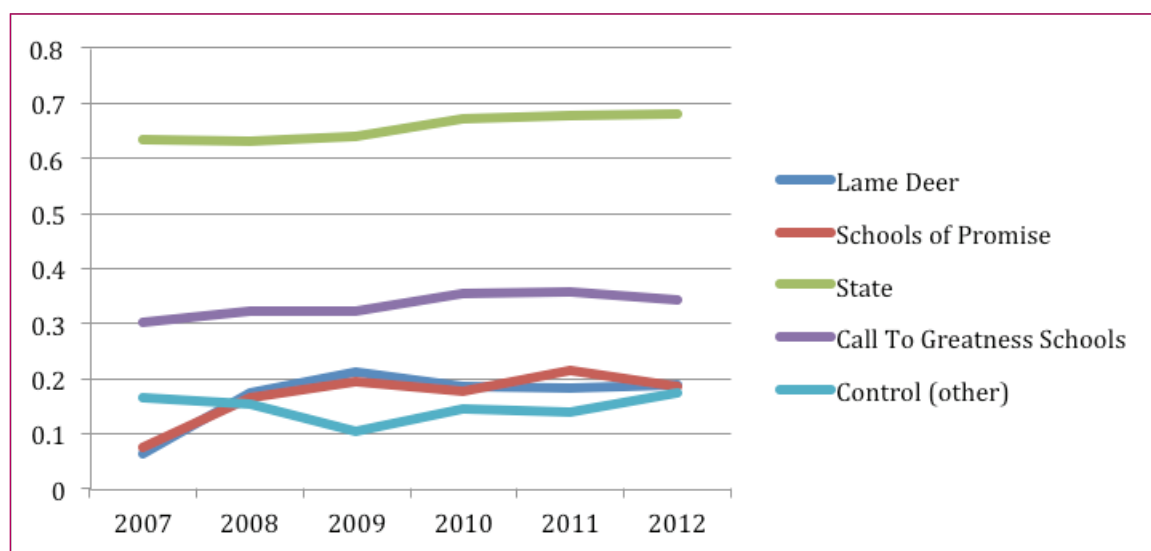
that the decline was a statistical outlier, and that the 2013 administration will be more aligned with the historic improvement trend. Nevertheless, even with the recent drop in performance, Frazer managed to maintain overall growth for the period of SIG administration, as indicated by Figure 22.

Figure 22. Mathematics growth by school groupings, 2009-2012



Had Frazer simply maintained its exam performance from 2011, it would have exceeded the improvement in all other relevant comparison frames. Given its actual performance, however, it managed to outpace the Schools of Promise trend.

Figure 23. Percent of students achieving proficient or advanced levels of performance on Mathematics assessments of the MontCAS, 2007-2012



Lame Deer's mathematics performance, summarized in both Figure 23 and Figure 24, shows the same struggle to break historic achievement patterns that appears to characterize other similar schools

– either in the CTG or SOP. More attention will need to be devoted to curricular and instructional interventions associated with mathematics in Lame Deer.

Figure 24. Mathematics growth by school groupings, 2009-2012

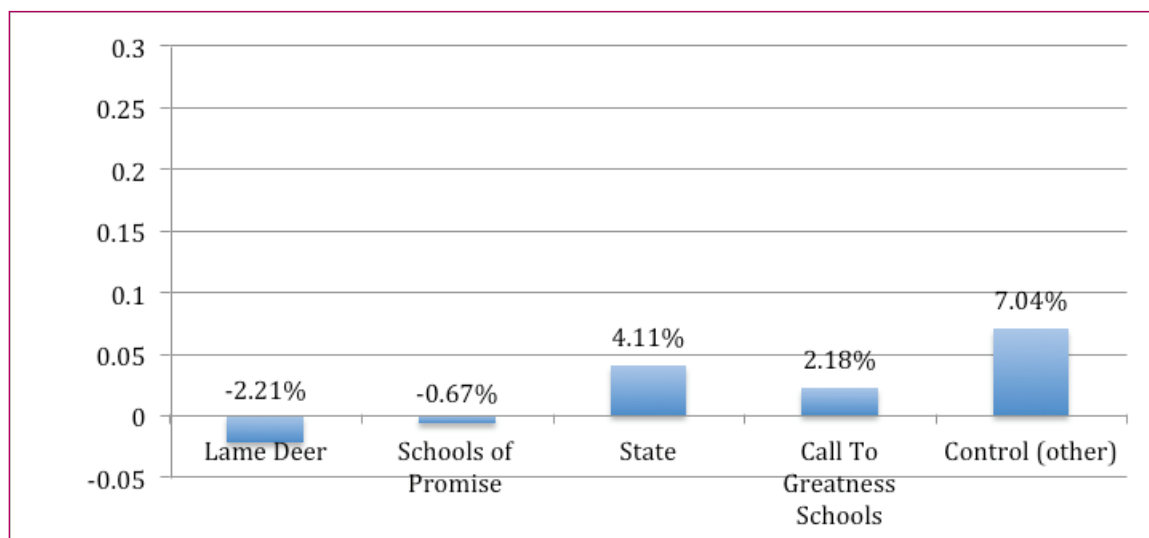
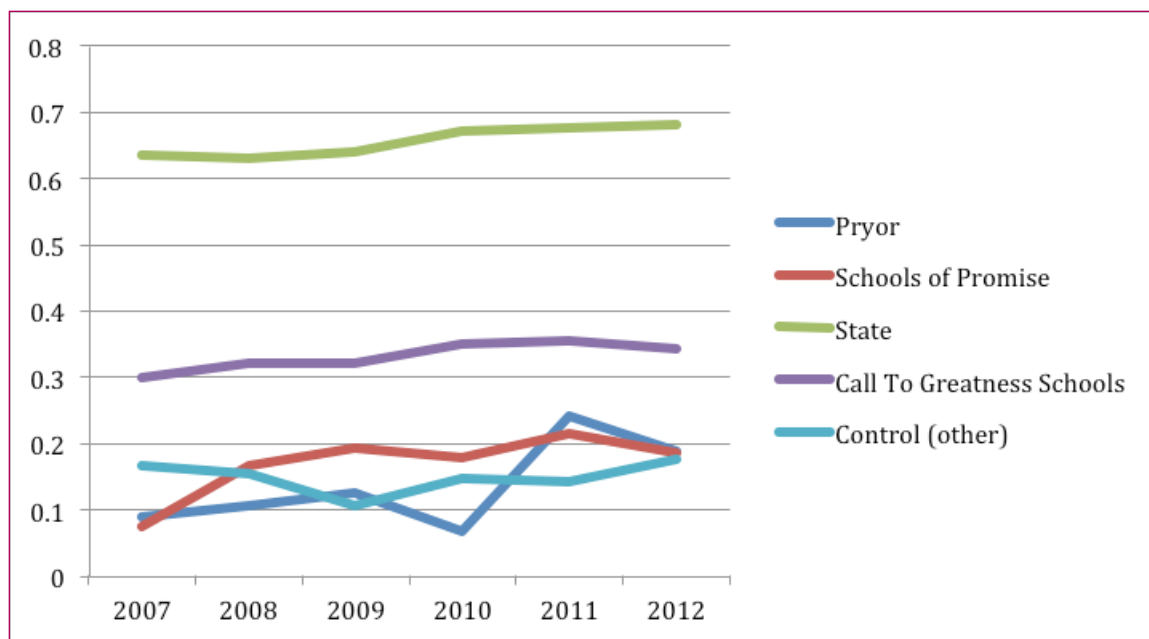
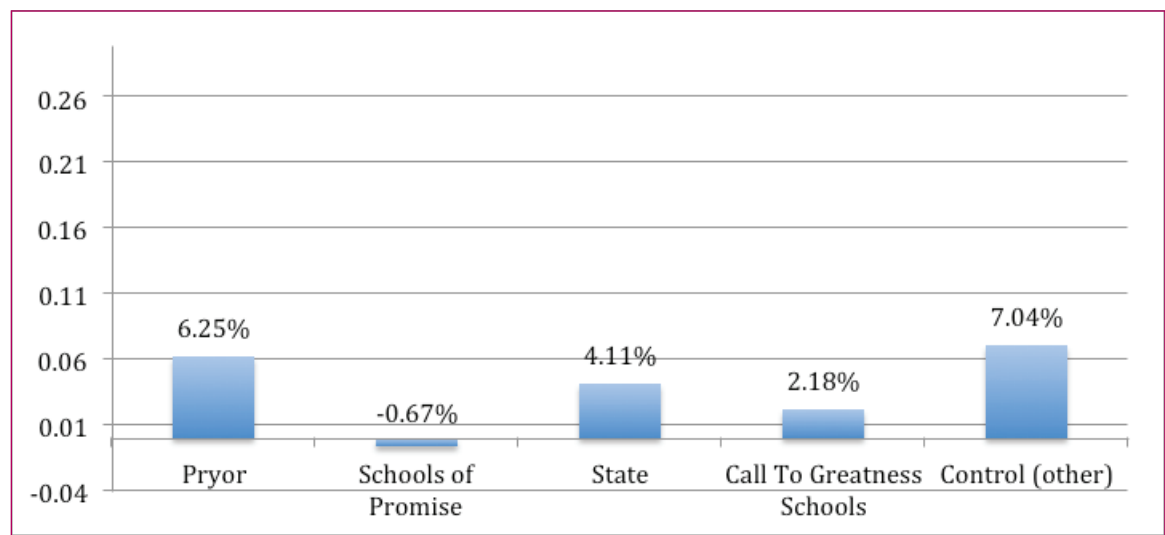


Figure 25. Percent of students achieving proficient or advanced levels of performance on Mathematics assessments of the MontCAS, 2007-2012



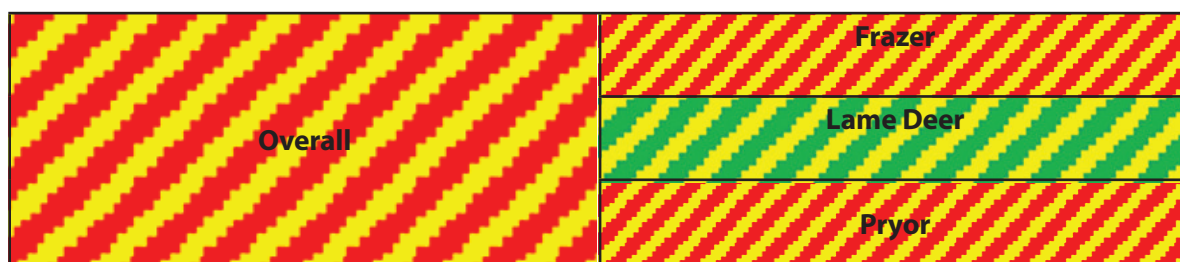
Pryor's mathematics achievement, as demonstrated by Figure 26 would seem to warrant a green rating, but the performance dynamics made evident in Figure 25, indicates a more complicated achievement profile – declining in both 2010 and 2012, with sharp increases in 2011. As a result, Pryor's performance is rated amber-green.

Figure 26. Mathematics growth by school groupings, 2009-2012



SCIENCE IMPROVEMENT

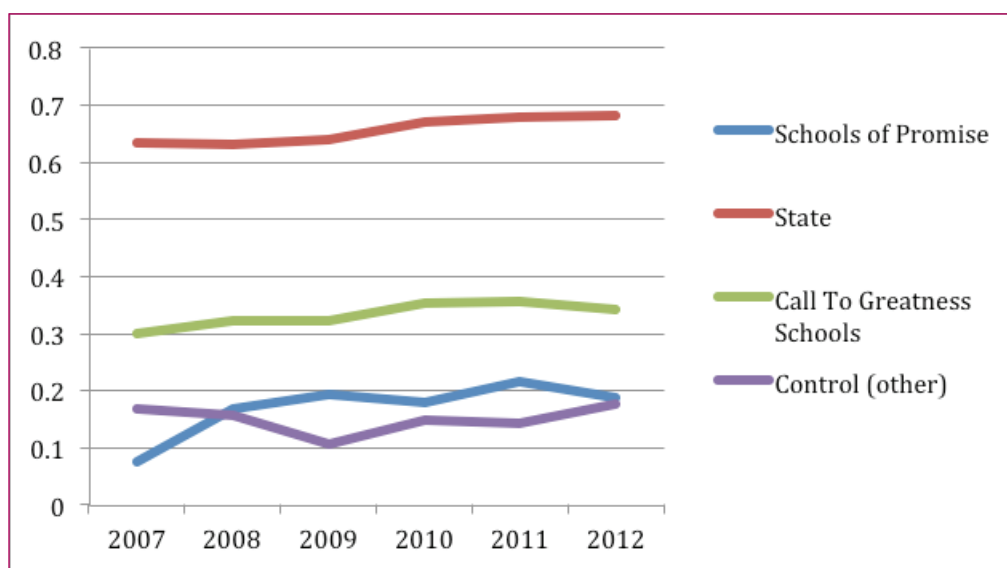
Figure 27. Science Assessment Improvement in the Montana Schools of Promise



Incidentally, the site level determinations for science were the inversion of the mathematics rating – Lame Deer performing at the amber green level, while Pryor and Frazer were ranked as amber red. As a result, due to the decline in the percentage of students performing at the novice level was not as sharp as with mathematics performance, the overall ranking was amber red.

The science assessment statements are based on the performance of students in grades 3-8 and grade 10 on the MontCAS assessment.

Figure 28. Percent of students achieving proficient or advanced levels of performance on Science assessments of the MontCAS, 2007-2012



A sharp spike in science scores in 2009 in the Schools of Promise quickly evaporated in the 2010 administration (see Figure 28). Since then, the scores have rebounded, outpacing the growth in the percentage of students achieving proficiency or higher in CTG Schools, in other schools initially designated as Schools of Promise that later withdrew from the program, noted diagrammatically as 'control (other)', and the state as a whole.

Figure 29. Science growth by school groupings, 2009-2012

As is visible in Figure 29, performance in the most recent test administration lagged slightly (not statistically significant lag) from 2009 levels.

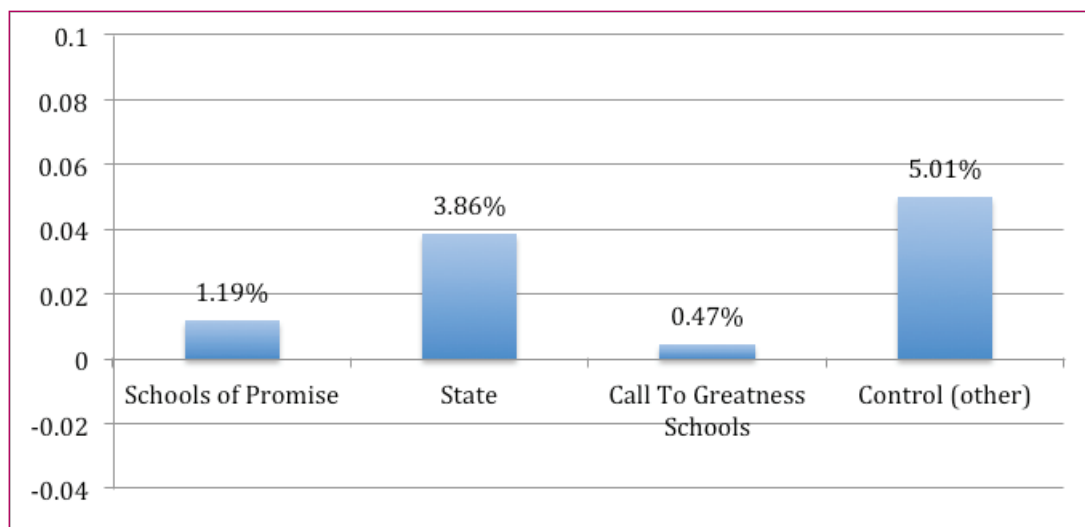
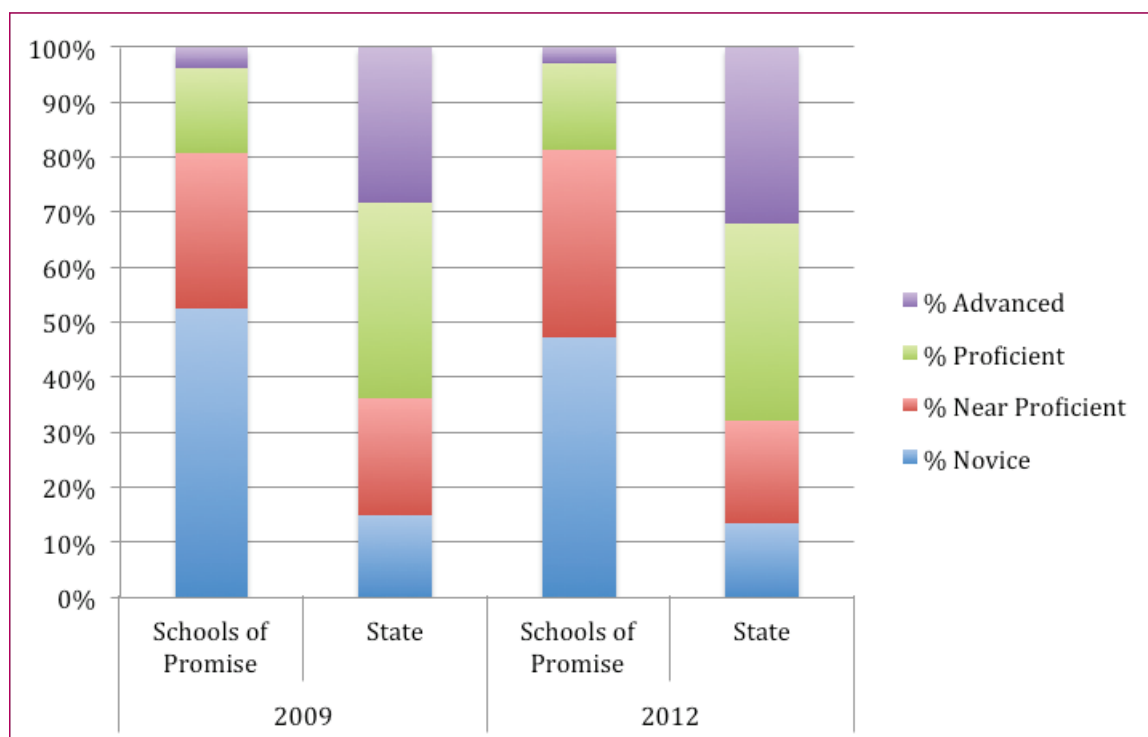


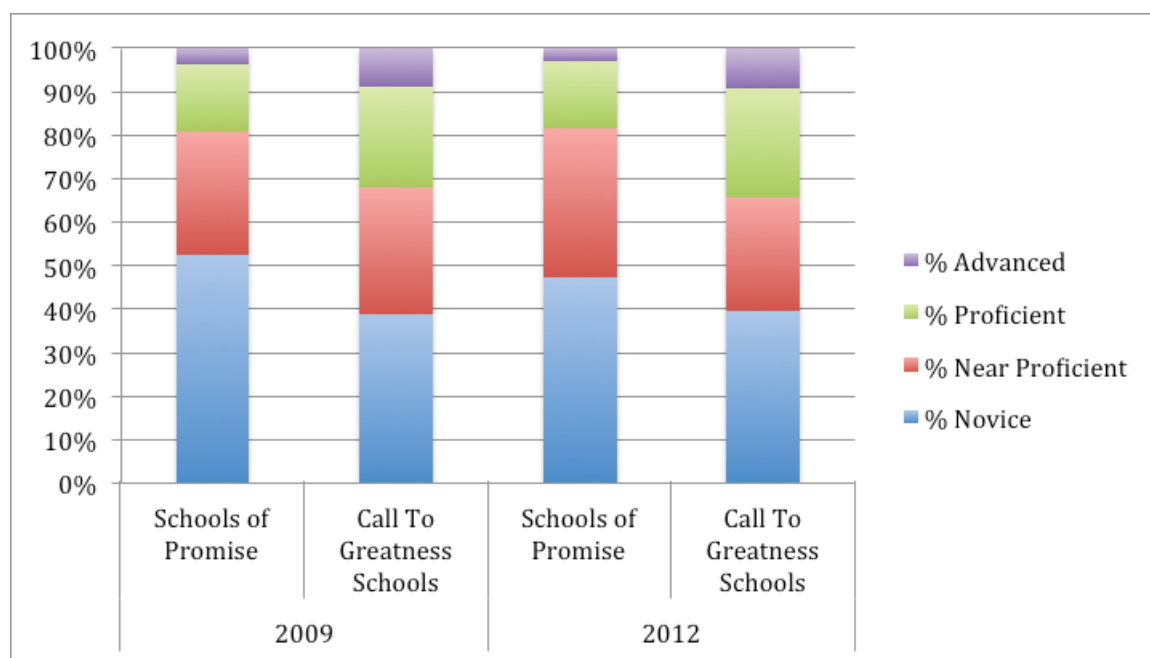
Figure 30. Performance distributions on the MontCAS Science Assessment, 2009 and 2012



The percentage of students who failed to score proficient on the state science exam fell by roughly 3% at the state level. By contrast, the Schools of Promise were only able to shrink the population of students failing to meet proficiency by roughly 1%. While this particular comparison alone would cast

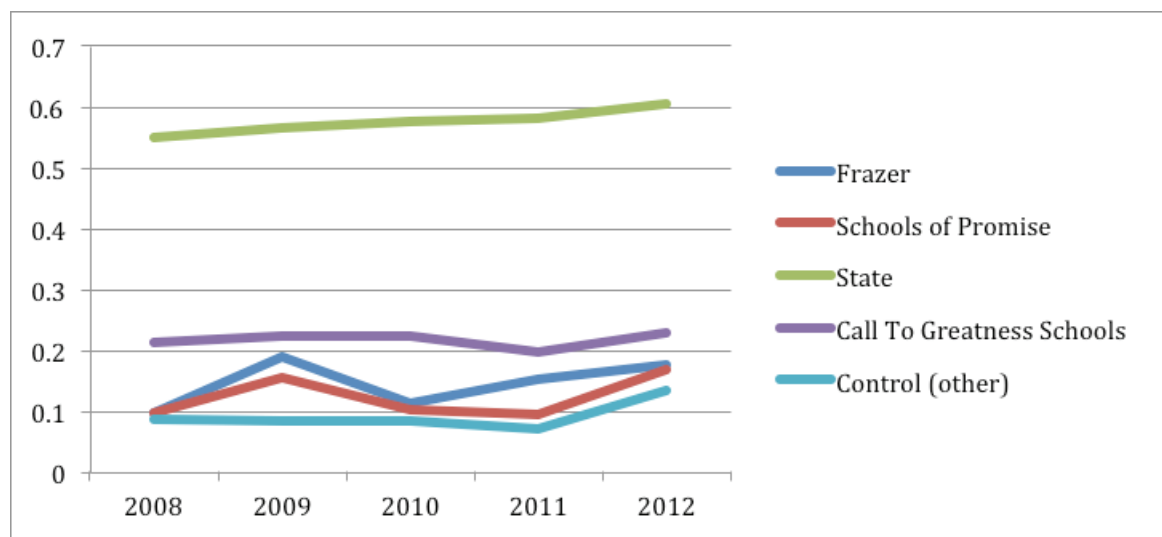
some doubt on the value of the Schools of Promise investment, it would also be important to note the rapid decline in the percentage of students scoring at the novice level. While the state percentage of novice test-takers fell only 1%, the Schools of Promise population shrunk about 5 to 6% over the time interval. Though still not quite meeting proficiency, the distribution of science scores is trending positively toward higher levels of performance. Publicly released reports will not detail distributions by school level, because the number of students scoring at any particular distribution is too small to protect statistical anonymity. Although notable, the magnitude of change in science scores was not as great as the change observed in mathematics, and does not merit an amber green determination overall.

Figure 31. Performance distributions on the MontCAS Science Assessment, 2009 and 2012



When comparing just to more demographically and historically-alike skills in Figure 31, there is reinforcement of a credible, positive, differential change in the Schools of Promise, where the decline in the percentage of novice students was roughly four times greater than in their comparison schools.

Figure 32. Percent of students achieving proficient or advanced levels of performance on Science assessments of the MontCAS, 2007-2012



Because Frazer was unable to outpace the growth in any of the comparison school groupings over the interval of grant administration (Figure 33), it would be difficult to rate performance as either green or amber green. Frazer avoids a red rating, though, because of promising trend performance (Figure 32), since 2012.

Figure 33. Science growth by school groupings, 2009-2012

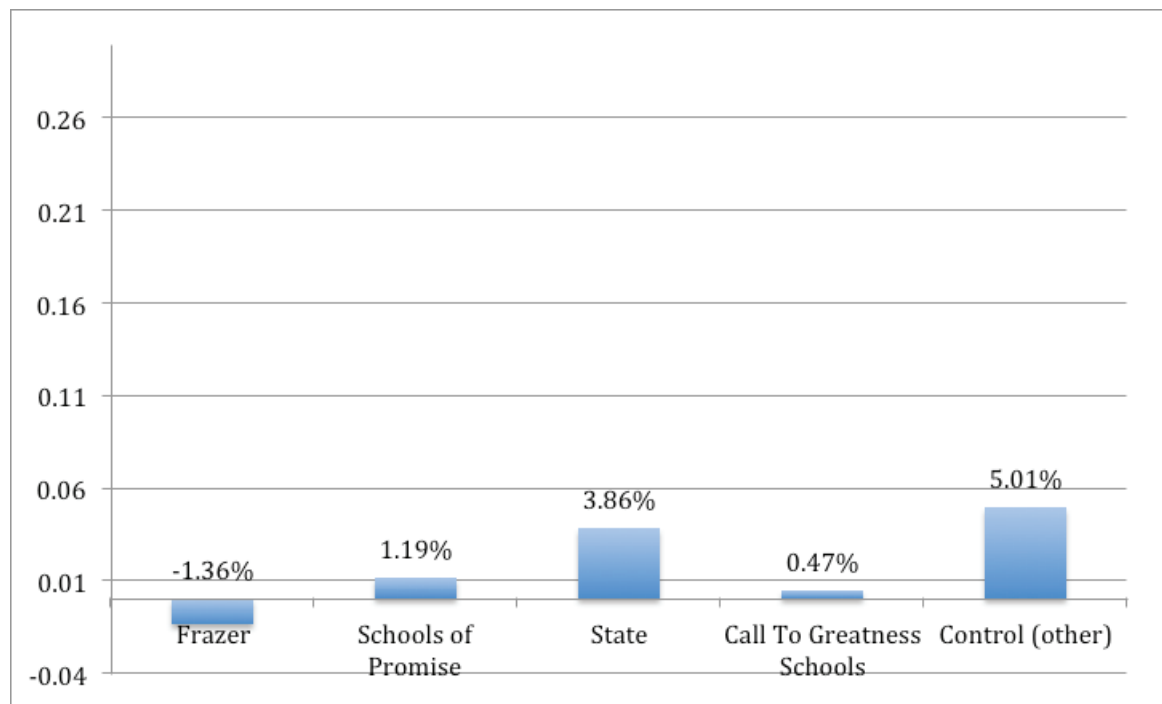
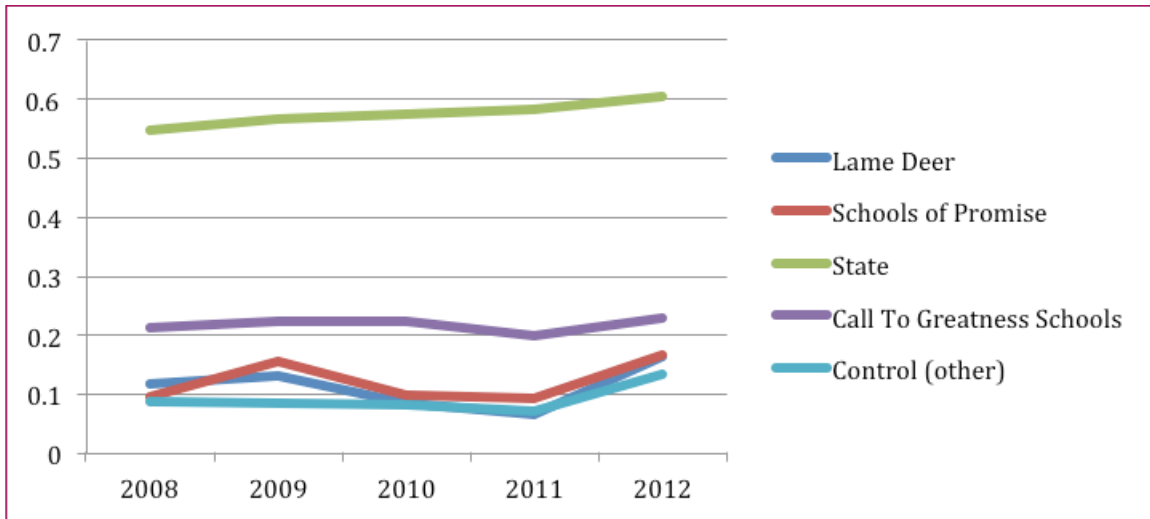


Figure 34. Percent of students achieving proficient or advanced levels of performance on Science assessments of the MontCAS, 2007-2012



Trend performance (Figure 34) and absolute growth rates over the grant administration interval (Figure 35) suggest an amber green determination for Lame Deer's science achievement. Overall growth exceeds at least one comparison group (the CTG schools), but still lags behind state growth, casting doubt on the efficacy of science interventions. Moreover, while there was overall improvement thanks to significant gains in science achievement last year (in fact, the steepest rate of improvement observed in any of the comparison groupings), it followed two years of decline. So while last year's growth, alone, would merit a green determination, there are enough reasons to be cautious about the growth that an amber green determination appears more prudent.

Figure 35. Science growth by school groupings, 2009-2012

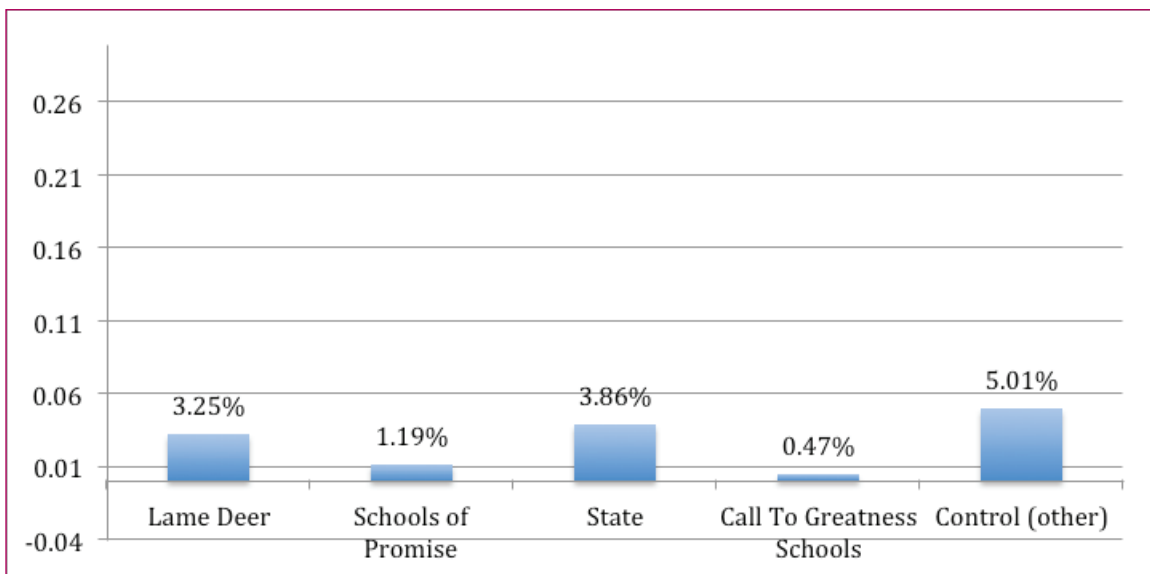
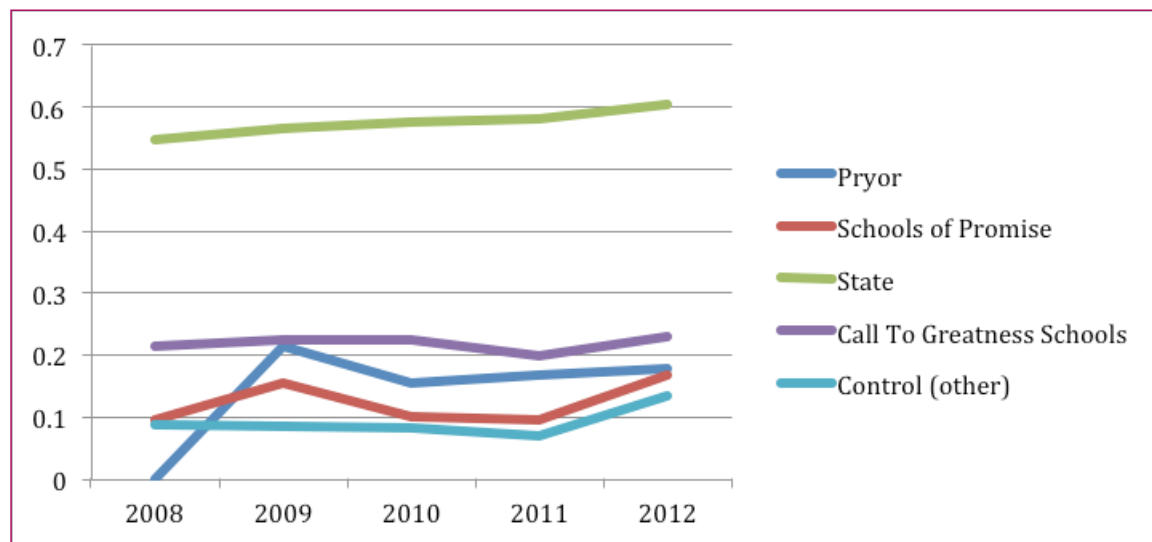
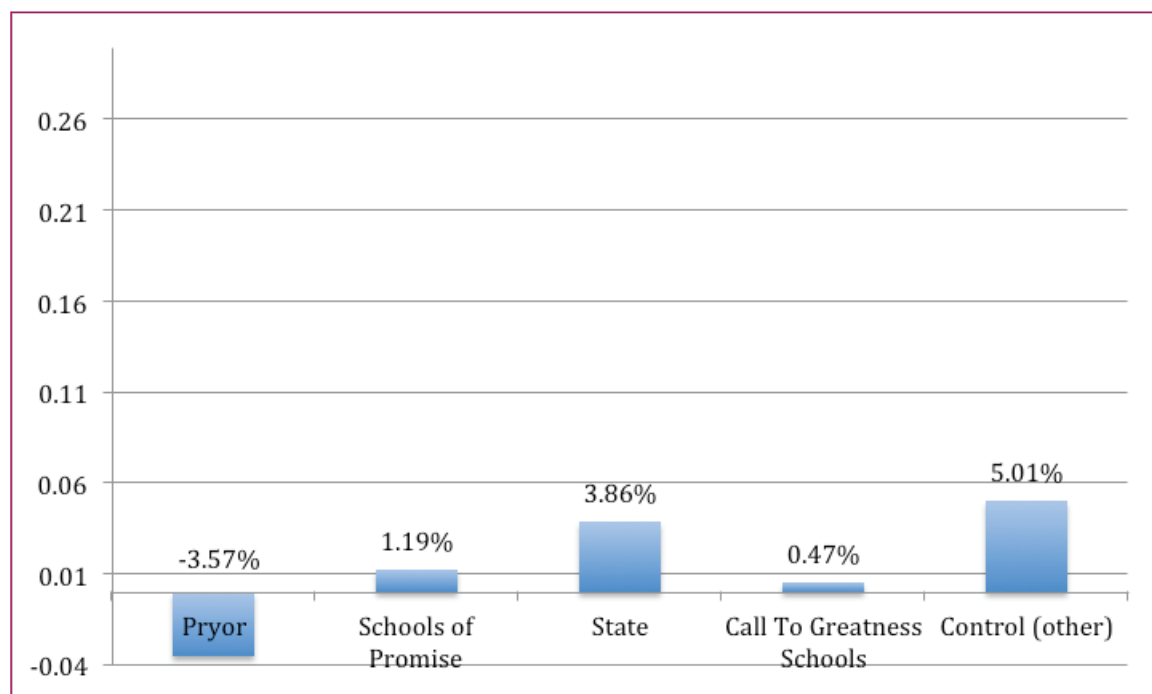


Figure 36. Percent of students achieving proficient or advanced levels of performance on Science assessments of the MontCAS, 2007-2012



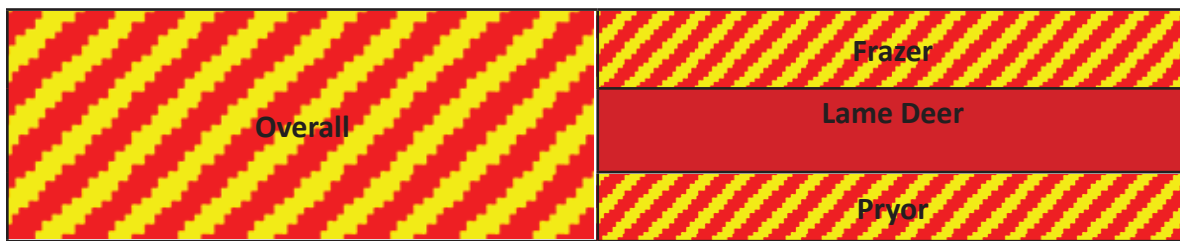
Pryor, who experienced a relatively sharp spike in science achievement in 2009 (Figure 36), gave a decent amount of those gains away in the very next test administration. Nevertheless, since the 2010 testing cycle, Pryor has experienced slow and steady science achievement growth that bodes well for system health, and merits an amber-red, rather than a red, rating.

Figure 37. Science growth by school groupings, 2009-2012



GRADUATION RATES

Figure 38. Graduation rate improvement in the Montana Schools of Promise



State Superintendent of Public Instruction Denise Juneau has made graduation rate performance a major focus of her administration. But while statewide improvement in graduation rates has been realized, similar gains have proven elusive for the Schools of Promise. Overall graduation rates have been erratic in the Schools of Promise for reasons still attempting to be understood by school-level practitioners (see Figure 39). One question focuses on the historic data quality – in other words, though rates appear to have fallen in some schools, some openly wonder whether this may be the first time that the counts have been properly conducted. The Montana OPI may wish to commission some research on the drivers of graduation rate performance in the SOP, given the centrality of graduation to eventual success for both individuals and communities, and given their importance in the State Superintendent’s strategic vision.

Figure 39. Schools of Promise Graduation Rates, 2008-2012

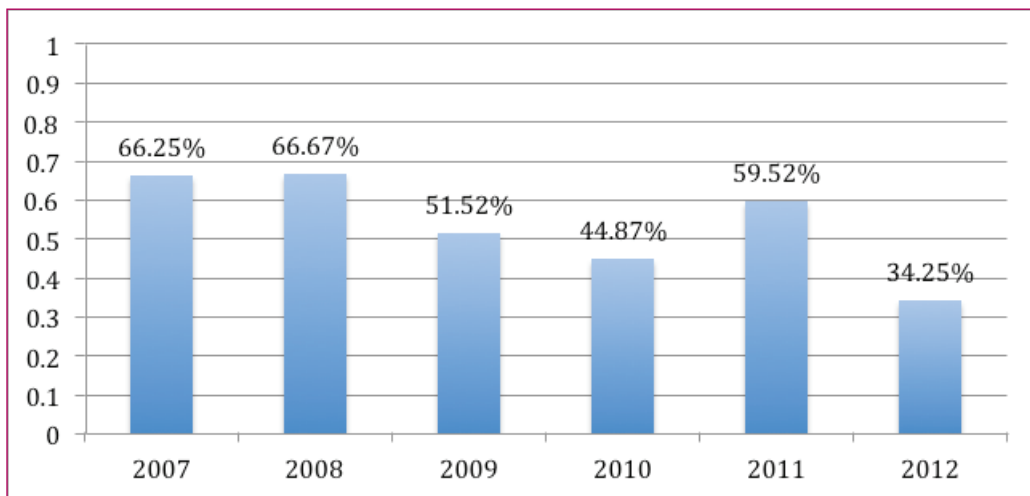
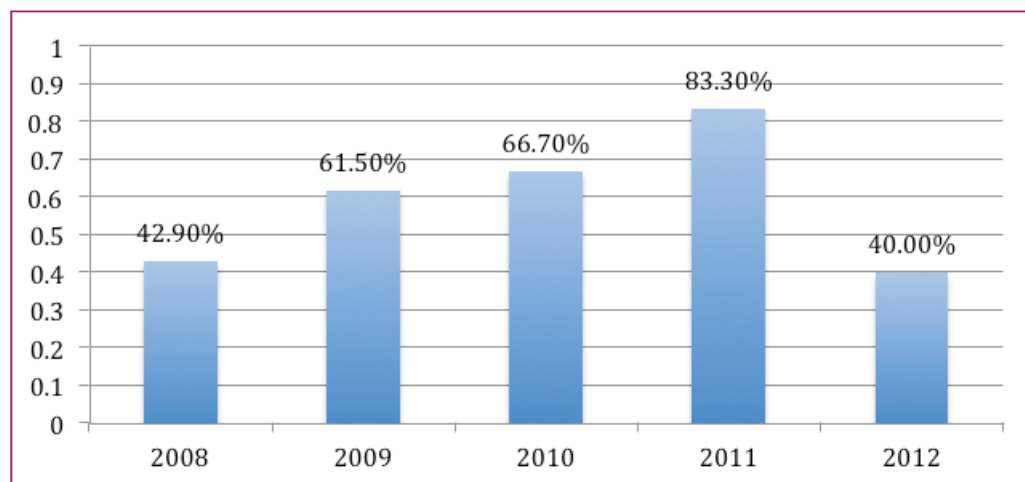
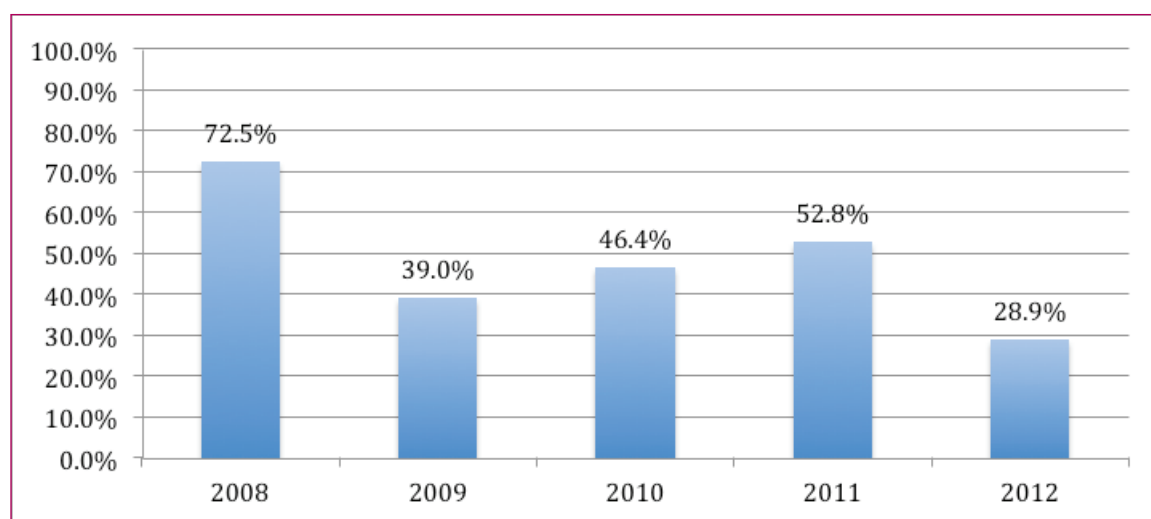


Figure 40. Frazer High School Graduation Rates, 2008-2012



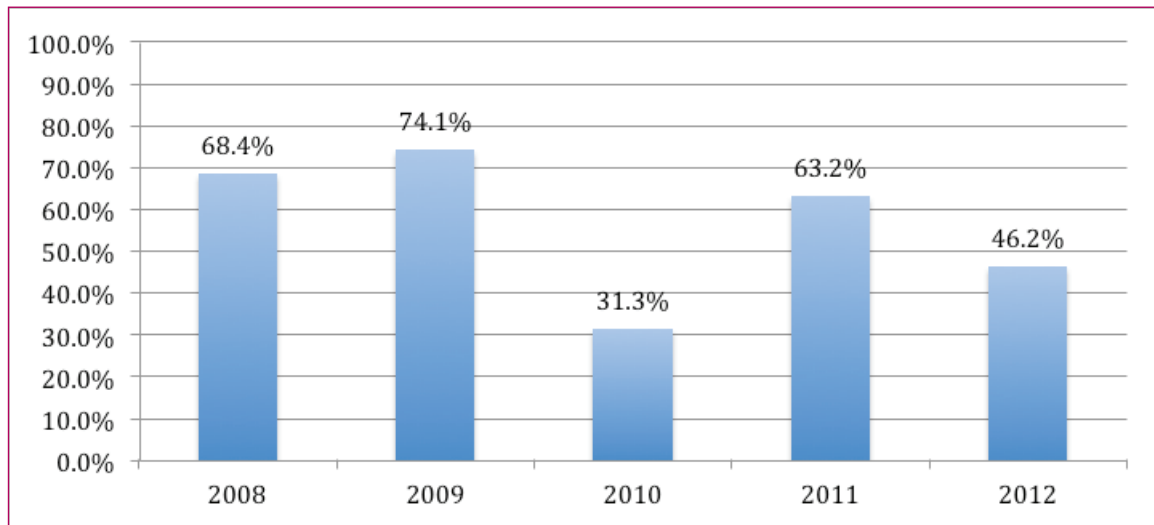
As shown in Figure 40, steady improvement typified Frazer’s graduation rate performance, putting it on track for a green rating. But a sharp decline in the 2012 graduation rate was enough to merit significant concern, and an amber red rating.

Figure 41. Lame Deer High School Graduation Rates, 2008-2012



Lame Deer also experienced steady growth over the period of the grant, though a significant drop in the percentage of graduates marked the first year (Figure 41). After two years spent reclaiming some of the 2008-2009 loss, the 2012 academic year gave way to a dispiriting decline in the percentage of graduates to just 28.9%. The data was so troubling, in fact, that Superintendent Juneau, joined by other OPI staff, traveled to Lame Deer in March 2013 to convene the community to talk about solutions to the dropout dilemma faced by Lame Deer. Given the decline, and the very low rate of graduation, the condition in Lame Deer warranted a red rating.

Figure 42. Pryor High School Graduation Rates, 2008-2012



Graduation rates at Pryor were inconsistent and, in two years, below 50%. As a result, the determination for Pryor was amber-red.

ATTENDANCE RATES

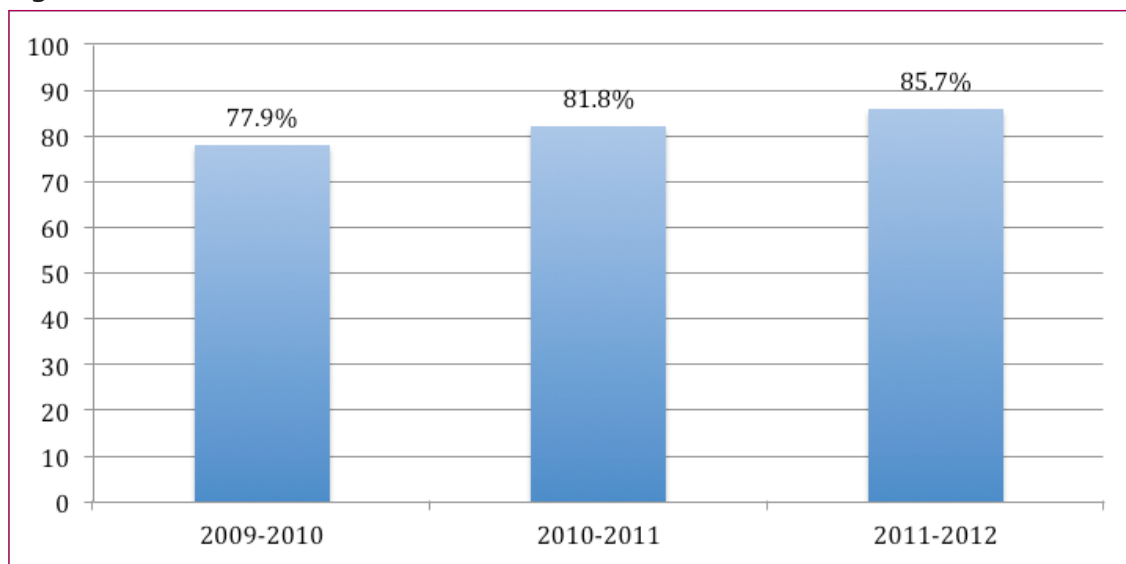
Figure 43. Overall Attendance improvement in the Montana Schools of Promise



In many ways, including attendance in the overall rubric contemplates the eventual use of the rubric. Presently, the value of looking at attendance data is quite marginal, as the attendance rates only reflect the attendance rate on two days of the academic year – one during the fall semester, and one during the spring semester. The attendance rates on those days are used for statutory purposes to calculate the Average Number Belonging (ANB) of a school, a key figure in determining state level funding.

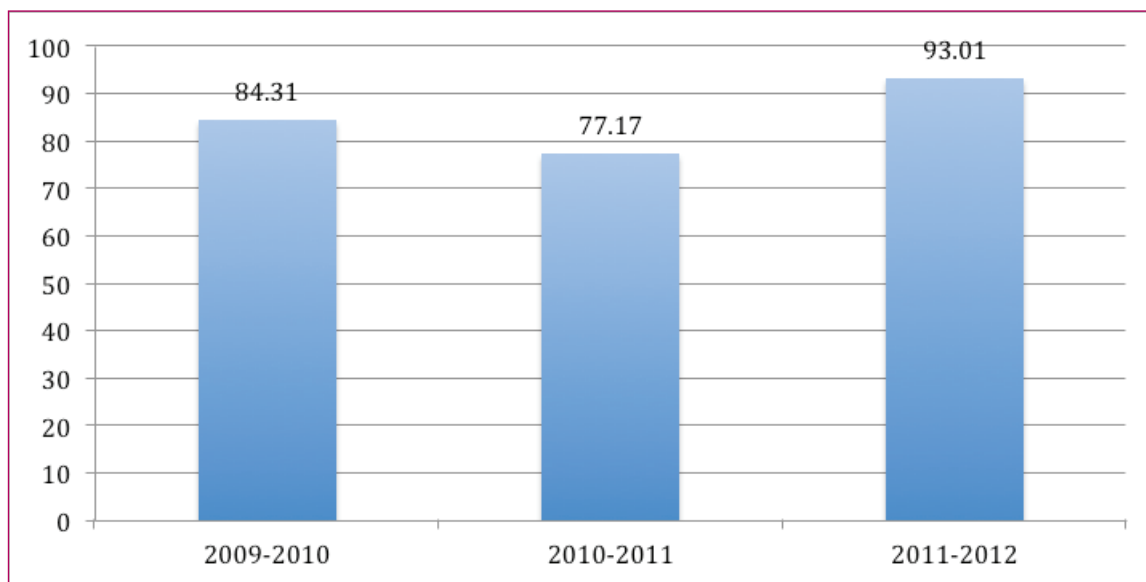
In the future, however, a more advanced data system procured by the state will allow for daily attendance reporting, making the data much more meaningful as a relevant outcome measure. Given the long-term importance of monitoring attendance data, the data is included in the rubric.⁴

Figure 44. Schools of Promise Attendance Rates Overall



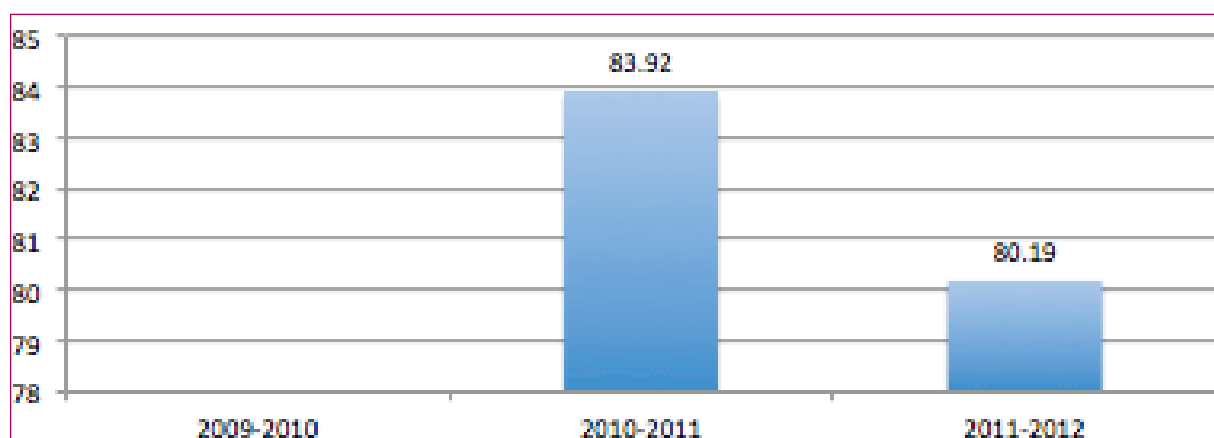
⁴ Normally when reporting data graphically across locations, it is important to make sure that data are reported against a common set of axes – for instance, from 0% to 100%. While that convention is adhered to throughout the present report, it was not followed in reporting attendance, suspension and expulsion rates. In the instance of attendance rates, it had to do with formatting challenges in the statistical software used to generate the graphs. With regard to suspension and expulsion data, it was because incidence rate values can often exceed 100%. Readers should be cautious in analyzing these sections for those reasons.

Figure 45. Frazer High School Attendance Rates, 2009-2012



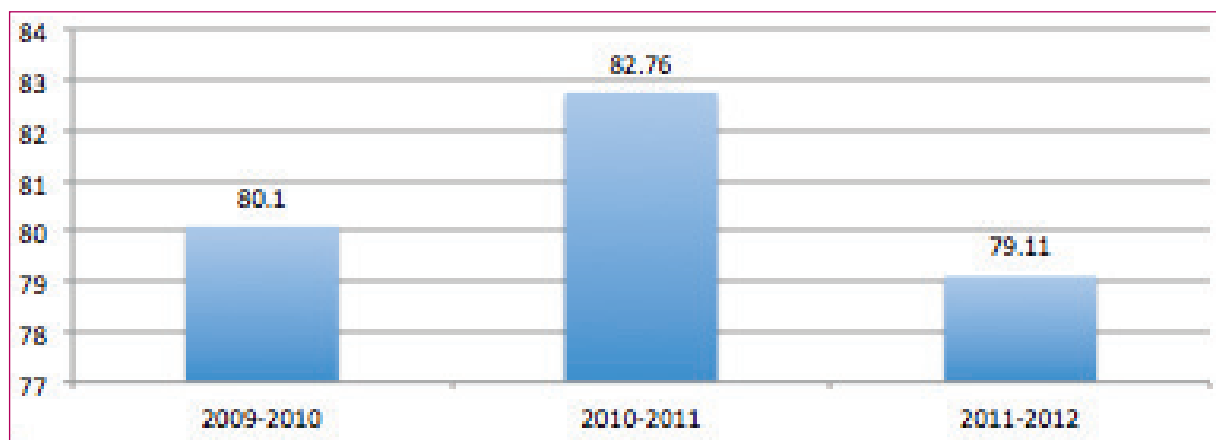
As shown in Figure 45, despite a decrease in outcomes in the 2010-2011 academic year at Frazer High School, the substantial gain (even when compared to the 2009-2010 baseline year) in 2011-2012 was enough to warrant an amber green rating. The unevenness of the trend cautioned against a green determination, however.

Figure 46. Lame Deer Middle School Attendance Rates, 2009-2012



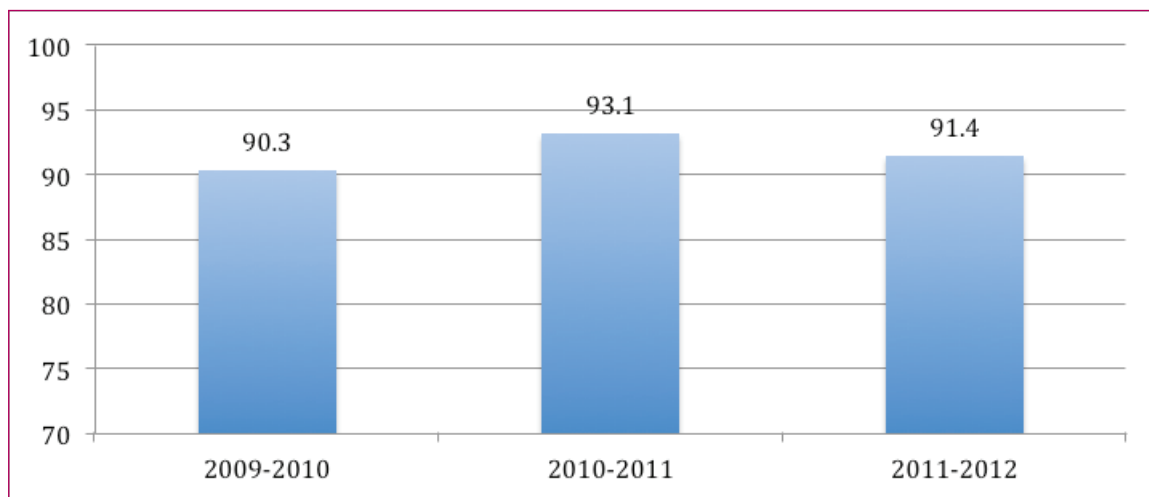
Results fell a statistically non-significant 3.72% over the two years measured at Lame Deer Middle School (see Figure 46). While the decline is non-significant, the lack of improvement is concerning.

Figure 47. Lame Deer High School Attendance Rates, 2009-2012



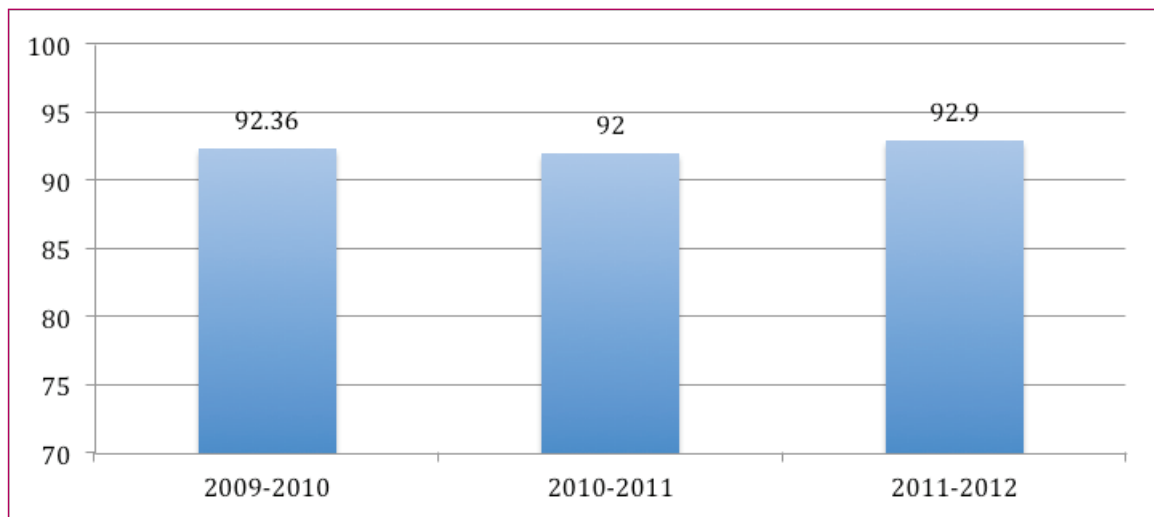
As shown in Figure 47, Lame Deer High School posted its lowest attendance rate in 2011-2012. Again, the variance was not statistically significant, but there are no signs of improvement at either the middle or high school level. Moreover, the stagnant rate is at a low level. If the rates are indicative of an average day (and likely, they overstate the average attendance rate), it means that 1 in every 5 students is out of school each day.

Figure 48. Pryor Elementary School Attendance Rates, 2009-2012



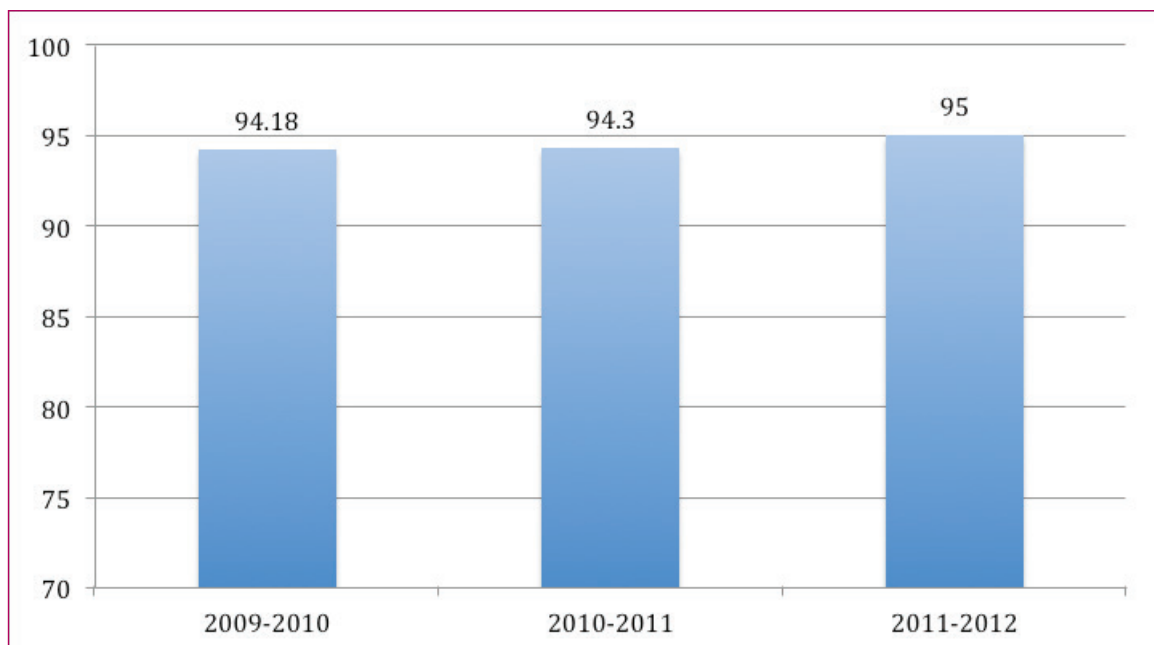
Like Lame Deer, there is no clear improvement in Pryor Elementary school attendance (Figure 48). Strangely, elementary school attendance is actually lower than middle school or high school attendance in Pryor. More should be done to understand the elementary school attendance rates, and what can be done to accelerate performance.

Figure 49. Pryor Middle School Attendance Rates, 2009-2012



Again, attendance rates at the middle school level appear stable, where improvement would be preferable.

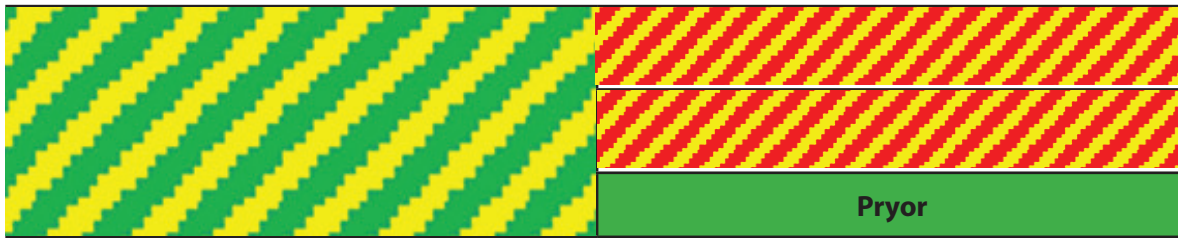
Figure 50. Plenty Coups High School Attendance Rates, 2009-2012



The high school attendance rates – while stable – are also relatively high. Contrasted with Lame Deer, where 1 in 5 students were absent daily, only 1 in 20 was missing in Plenty Coups High.

SUSPENSION RATES

Figure 51. Overall Suspension Improvement at Montana Schools of Promise



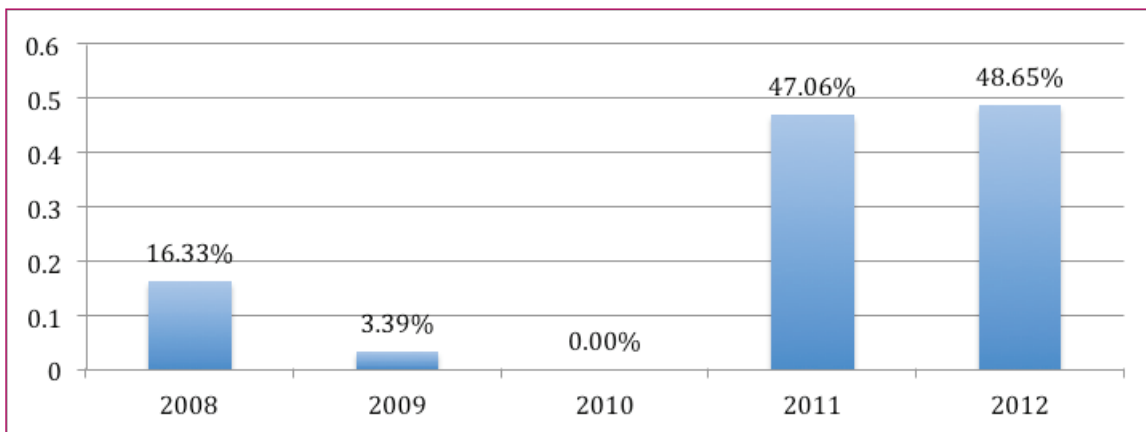
Suspension rates in the SOP demonstrate remarkable and difficult to understand variance both over time and across school locations, calling into question the reliability of the data. Treating the data as accurate, however, the overall trend appears to be slightly positive for the SOP as a whole.

It should be noted that the suspension and expulsion rates are reported as percentages of the overall school-going population. In other words, the formula used to calculate the rate was as follows:

$$\frac{\text{Number of suspensions (or expulsions, see below)}}{\text{Number of students enrolled}} = \text{Rate}$$

It is a popular misunderstanding, when reviewing such incidence rates, to view rates greater than 1 to indicate that all students were at one point in time suspended. Indeed, though, an incidence rate greater than 1 may indicate that few students were suspended several times.

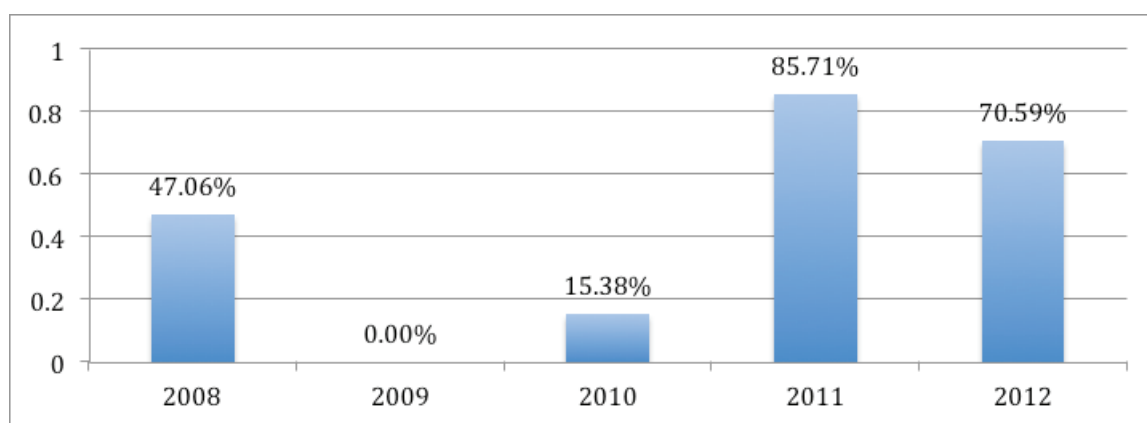
Figure 52. Frazer Elementary Suspension Rates, 2008-2012



As shown in Figure 52, elementary suspension rates appear to be heading in a dangerous direction in Frazer, perhaps warranting a red determination. It is a trend mirrored in both the middle school and the high school, though in the instances of the middle and high school (see Figures 53 and 54) the rate of suspension fell in the 2012 school year, suggesting some improvement. Moreover, in speaking with

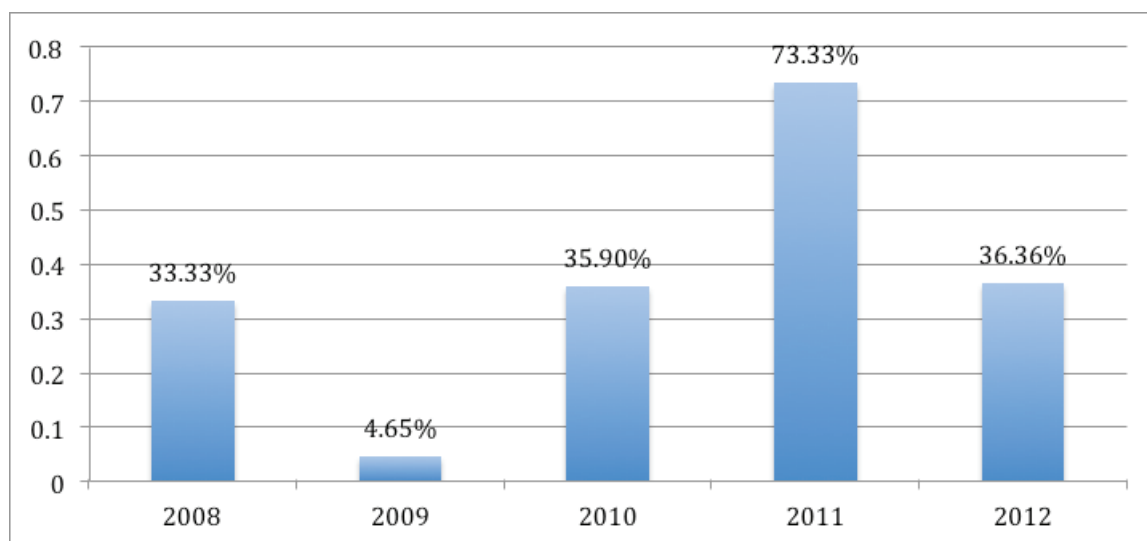
school representatives, there are mitigating factors around the data that caution against a view that there has been a sudden rash of suspensions. First, more than one educator suggested that the reason there may be an increase is that the school now has clearer behavior standards, a result of implementing the state's Montana Behavioral Initiative (MBI) standards. And the steady to declining rate suggests that they're being implemented, and that students are responding with better behaviors. Secondly, a competing or complementary narrative is that the historic data were just terribly inaccurate – that no one was rigorously counting or tracking the number of suspensions that actually occurred, thus artificially lowering historic rates. For these reasons, a determination of amber-red was made.

Figure 53. Frazer Middle School Suspension Rates, 2008-2012



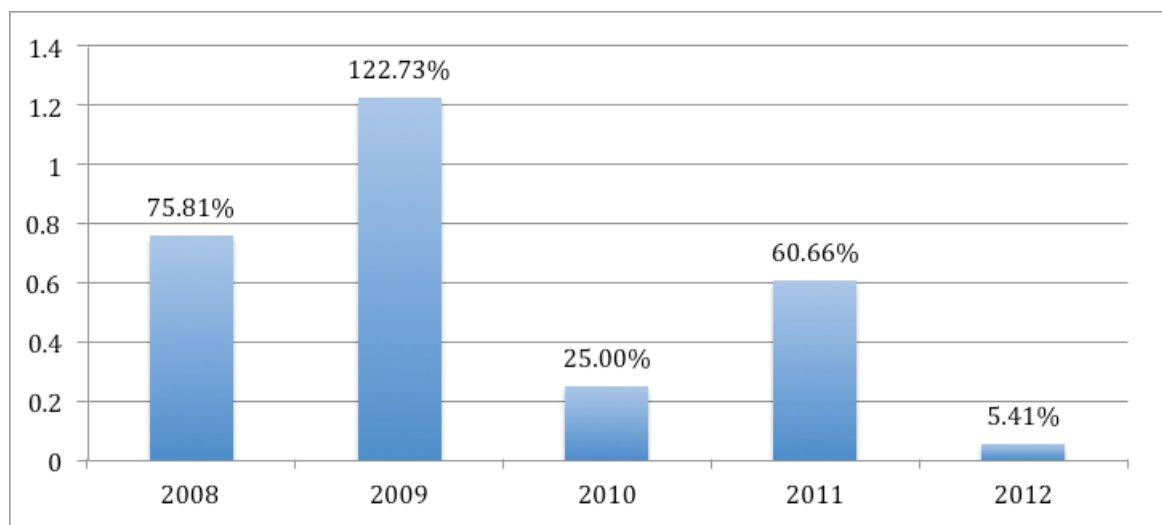
Middle school suspension rates can be understood in the same ways that elementary school suspension rates can be rationalized in Frazer (please see paragraph above).

Figure 54. Frazer High School Suspension Rates, 2008-2012



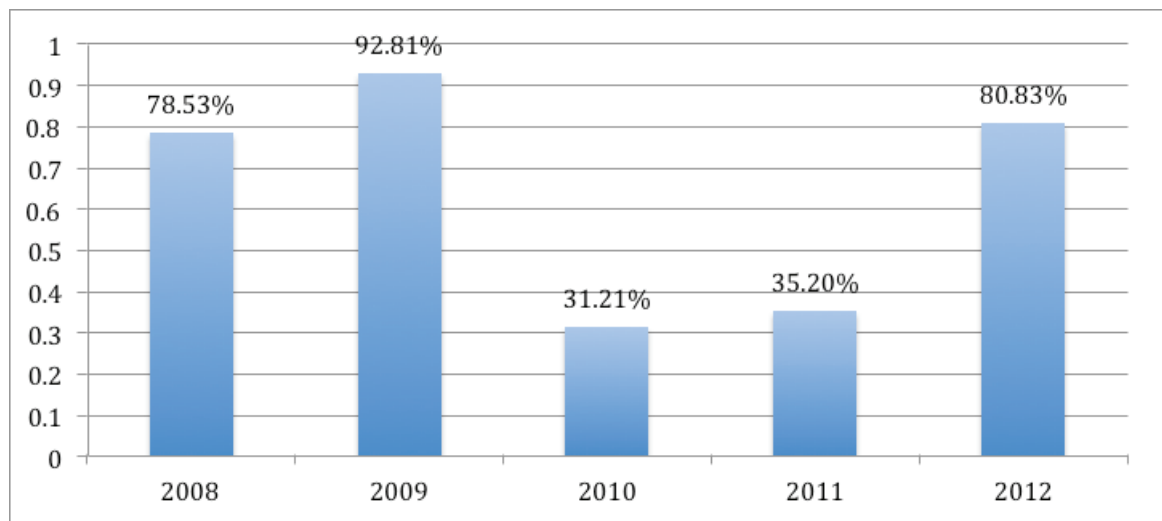
High school suspension rates can be understood in the same ways that elementary and middle school suspension rates can be rationalized in Frazer.

Figure 55. Lame Deer Grades 7-8 Suspension Rates, 2008-2012



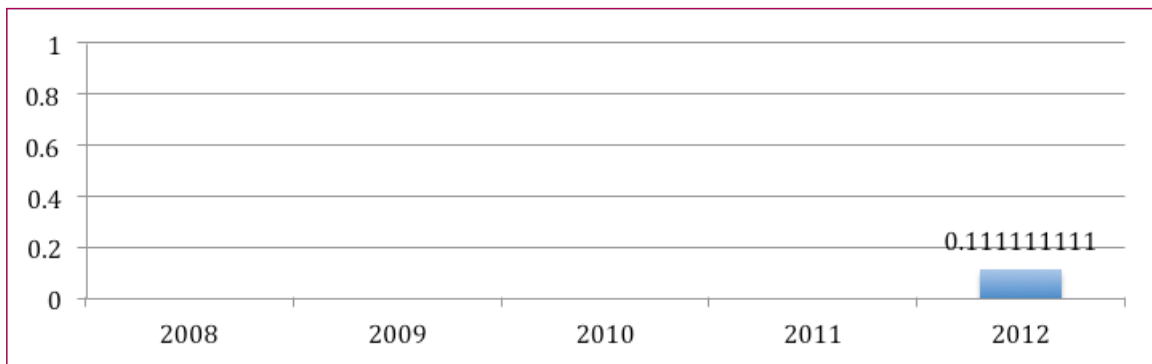
The Lame Deer Grades 7-8 rates show a generalized decrease over the term of the grant, warranting a green or amber green rating (see Figure 55). The rating however, needs to be weighed against the outcomes at the High School, which are not nearly as promising.

Figure 56. Lame Deer High School Suspension Rates, 2008-2012



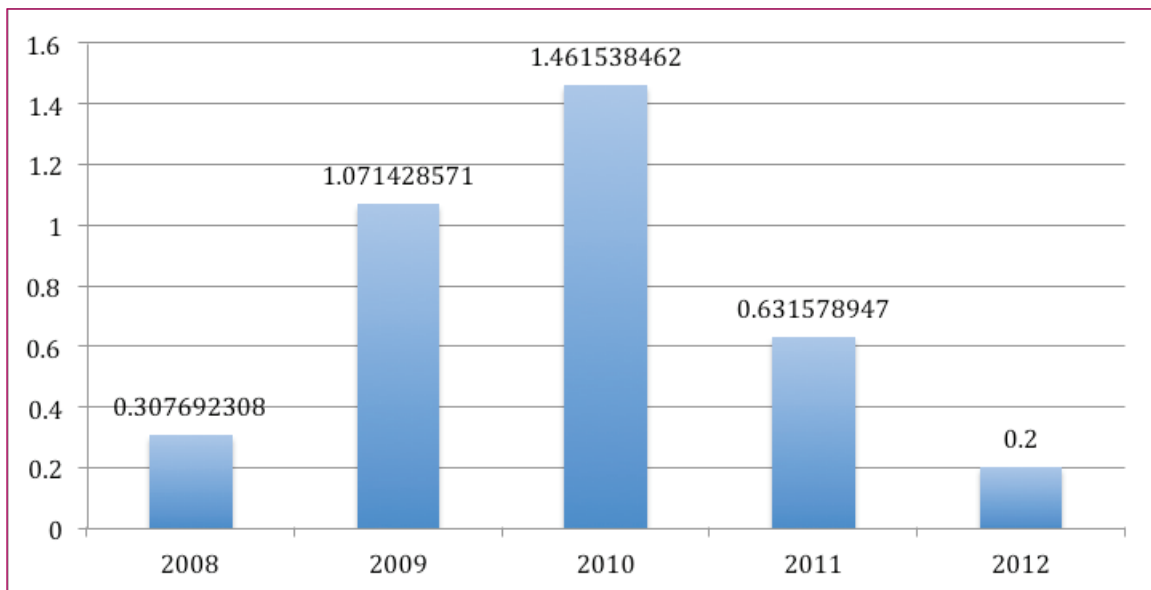
In Figure 56, the trend for the grant term is an increased reliance on suspension for behavioral corrections, warranting an amber red or red rating. Brought together, an overall amber red rating was determined because of the intuitions of those at the school that this reliance on suspension as a behavioral remedy may be linked to the high dropout rates observed in Lame Deer. Because addressing the graduation rate is such a high priority for the community, an amber red rating was assigned, to indicate the relative importance of addressing the suspension rate.

Figure 57. Pryor Elementary Suspension Rates, 2008-2012



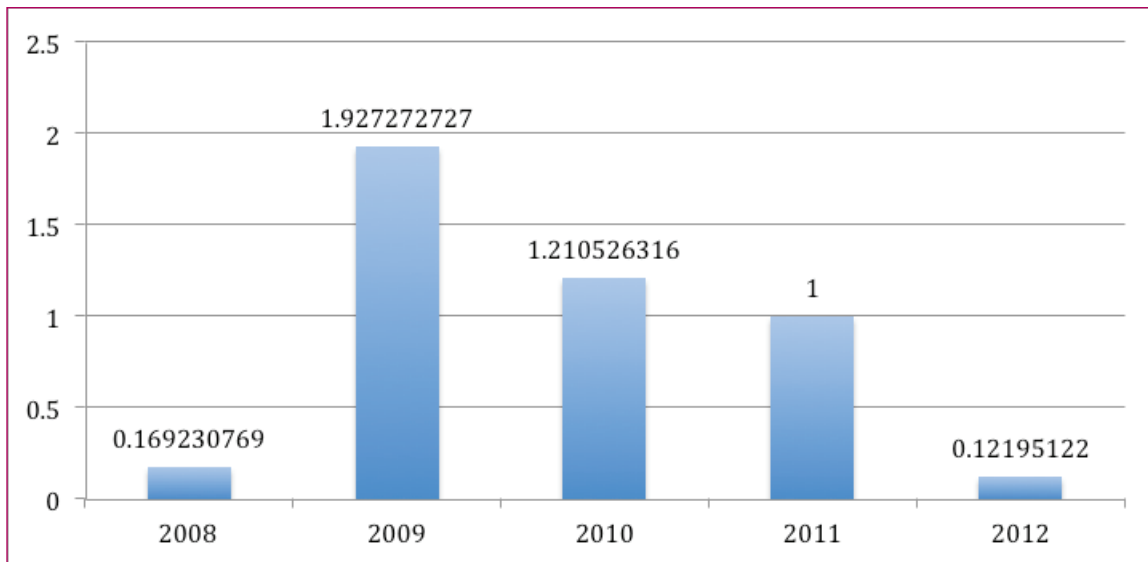
Unlike other school settings, officials in Pryor were relatively confident in their suspension data from the elementary schools. Out of school suspension is infrequently used as a remedy in Pryor elementary schools, and in 2012 there were isolated suspension incidents that did not seem to be systemically influenced, but rather isolated instances (see Figure 57).

Figure 58. Pryor Grades 7-8 Suspension Rates, 2008-2012



Unlike the elementary school, Pryor middle school has a historic memory of strong reliance on suspension as a disciplinary vehicle as shown in Figure 58. That reliance, however, is clearly waning, warranting a green or amber green rating.

Figure 59. Plenty Coups High School Suspension Rates, 2008-2012



Like the middle school in Pryor, Plenty Coups High School is witnessing a tremendous decline in suspension rates. Taken altogether, discipline seems much better controlled in Pryor than it was previously, warranting a green overall rating.

EXPULSION RATES

Figure 60. Overall Expulsion Improvement in the Montana Schools of Promise

Overall	Frazer
	Lame Deer
	Pryor

Like the suspension data, stakeholders in the SOP improvement process lack confidence in the quality and reliability of the expulsion data. There are serious questions about whether historic data, indicating few if any expulsions, is correct, as many stakeholders recall instances of expulsion. Nevertheless, they are certain about the present year's data, where zero expulsions were reported. Thus, results for all school sites were rated green, as were the overall SOP results.

Frazer Elementary School Expulsion Rates, 2008-2011

None reported

Figure 61. Frazer Grades 7-8 School Expulsion Rates, 2008-2011

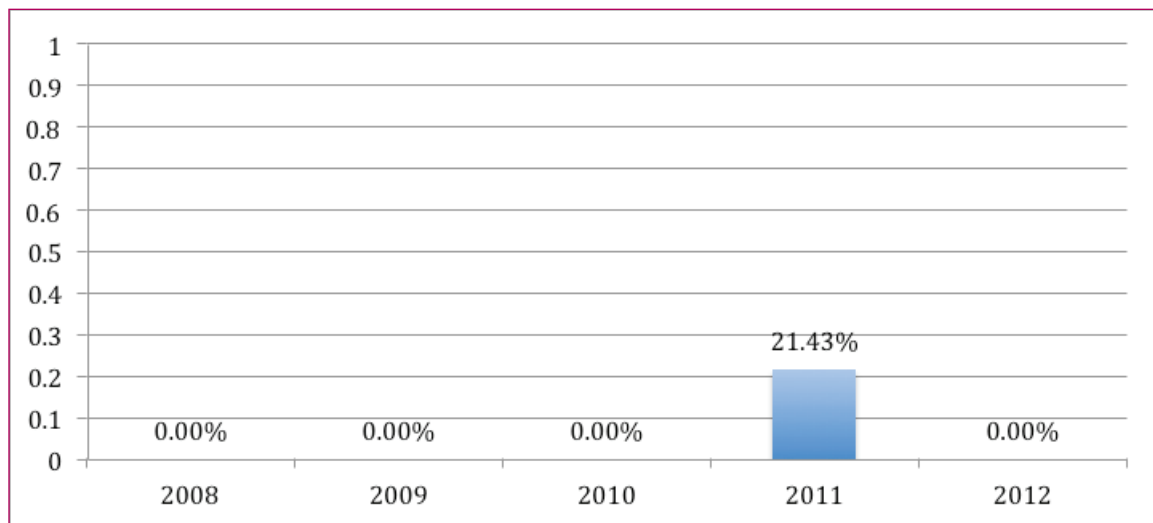
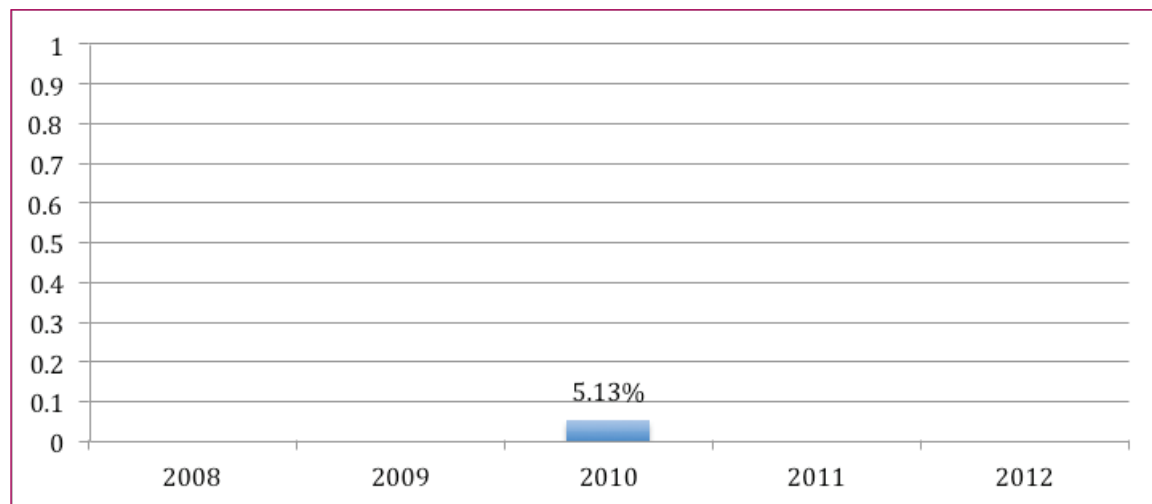


Figure 62. Frazer High School Expulsion Rates, 2008-2011



Lame Deer Grades 7-8 Expulsion Rates, 2008-2011

None reported

Lame Deer High School Expulsion Rates, 2008-2011

None reported

Pryor Elementary Expulsion Rate, 2008-2011

None reported

Figure 63. Pryor Grades 7-8 Expulsion Rates, 2008-2011



Figure 64. Plenty Coups High School Expulsion Rates, 2008-2011



PART III. PERCEPTION DATA AND FINDINGS

Several stakeholder groups were interviewed at the school sites in the spring of 2012. There was a certain set of guiding questions to help determine the efficacy of SOP interventions that were deployed at each site, with each group. The guiding questions – essentially the interview frameworks – are included as an appendix to the present report (Appendix B). But the interviews were meant to feel collegial and conversational, and so significant latitude was taken in the actual conversations, while still being mindful to address the substance of the interview framework.

For each stakeholder group, roughly 4-8 participants were invited to engage in conversation. Sometimes, as schedules allowed, individual meetings were arranged, and occasionally meetings with larger groups were placed on the calendar. Notes were compiled, and word clouds of each stakeholder groups' comments were generated.

In addition, a modified technique from complexity science was deployed to minimize the level of researcher bias in determining the overall impressions of the group. To do so, attendees of a SOP conference in January of 2012 were asked to engage in a scenario planning exercise. In the exercise, participants were asked first to craft several descriptions of the present state of their school. Participants wrote some reflections and descriptions on small notes of paper and clustered them.

Participants were then asked to describe the attributes of a "heaven" or "utopia" state. Again, they followed the same procedure as with their present-state descriptions.

Then they were asked to describe a path that would link the two patches of descriptions – what would have to happen in order to cause the present state to meet the ideal state? Participants described these paths, and linked their current state to the ideal.

The process was repeated again. In the scenario planning, participants were also asked to describe "hell" or "dystopia." Again, they did, and again, they drew links from where they were to a description of dystopia. The graph of the conversations is included below. The first graph visually represents the finished product, while the second labels some features of the product that would be understood by participants, given the development process.⁵

⁵ This technique is an adaptation of techniques used in complexity science, and attributed to researchers and complexity scientists within the Cognitive Edge Network. More about the technique can be found here: <http://cognitive-edge.com/library/methods/the-future-backwards-basic/>.

Figure 65. Results of scenario planning exercise with Schools of Promise, January 2012

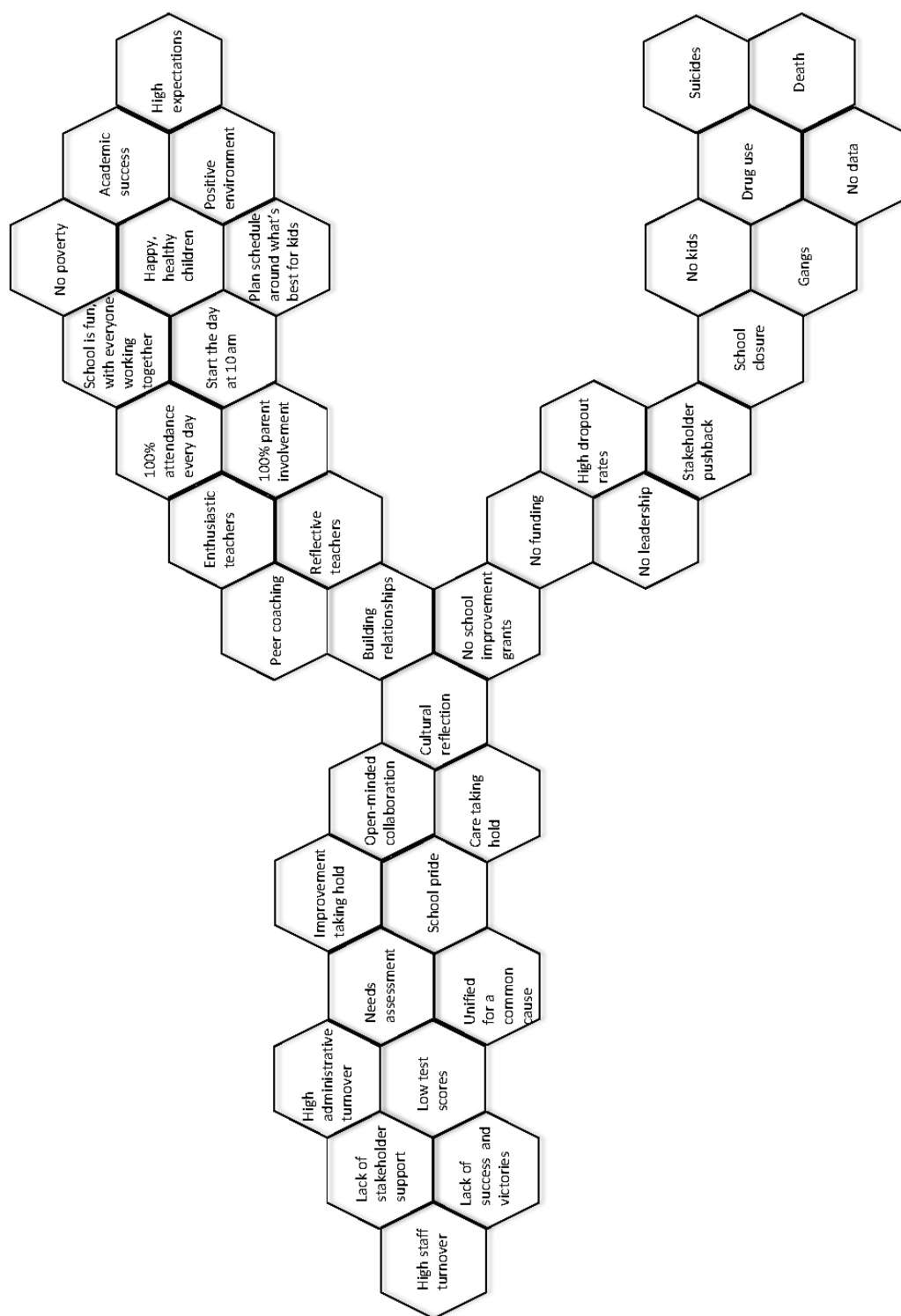
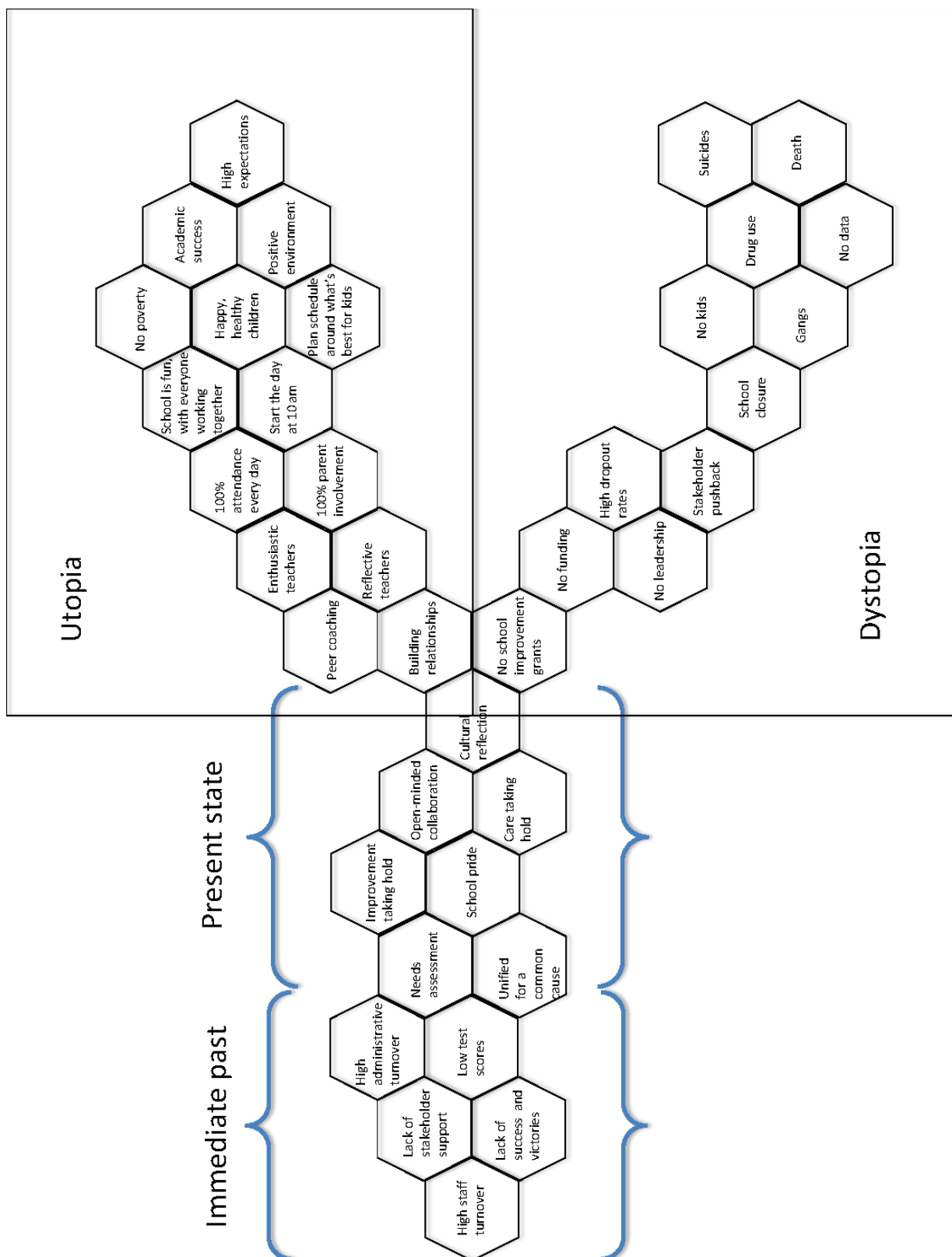


Figure 66. Results of scenario planning exercise with Schools of Promise, January 2012, with labels

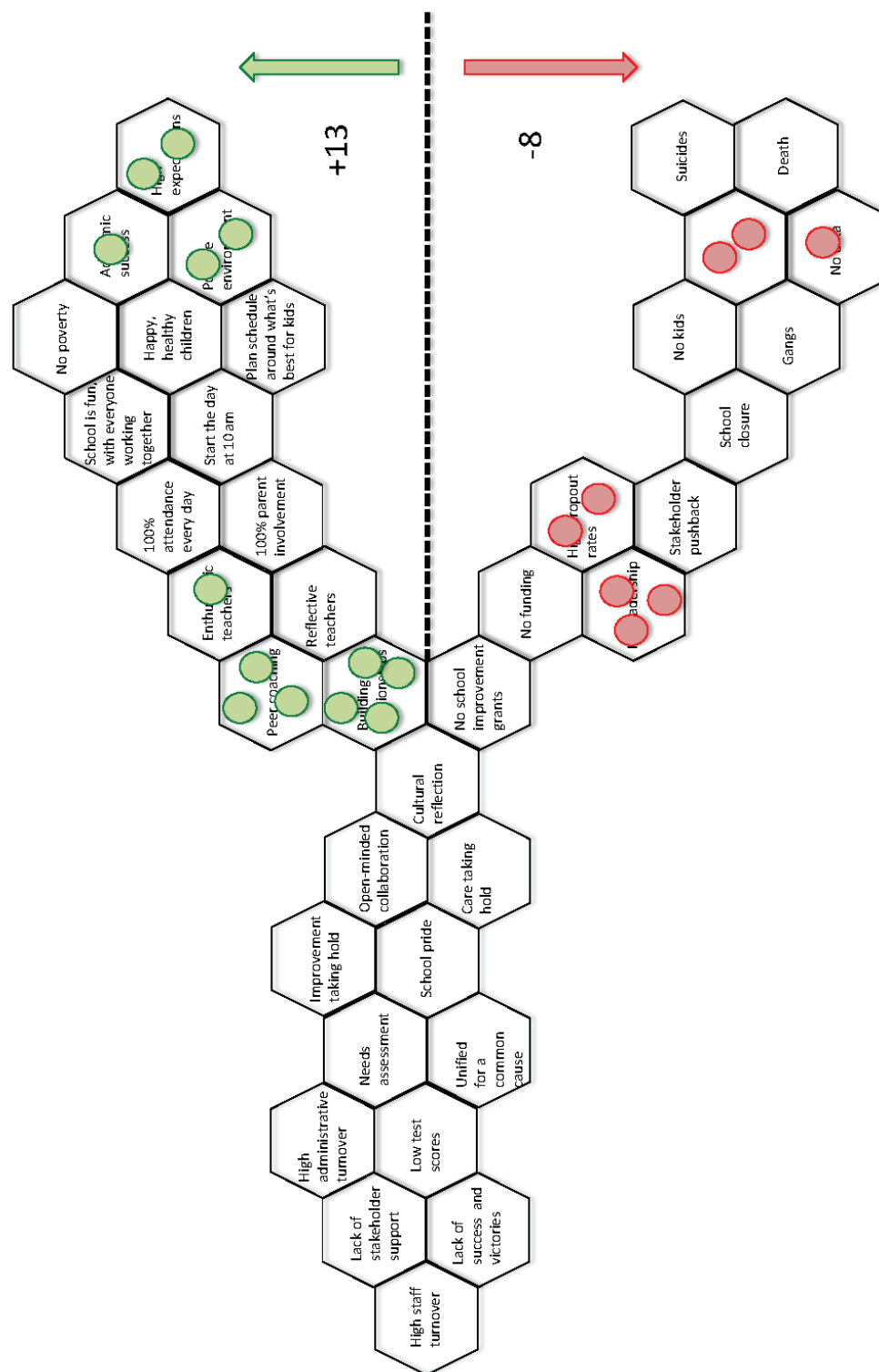


As a result, the descriptions of good and bad are not those of the researcher; they are instead those of the schools and communities themselves.

When interviews were conducted several months later, a look for stories and narratives that matched the descriptions from the scenario planning was undertaken. Those stories were either on a “path to utopia,” or a “path to dystopia.” True to the techniques of complexity science, participants were actually asked to show where their stories were on the scenario-planning map. Doing so minimizes another potential place for the insertion of researcher bias – attempting to say that a particular piece of qualitative feedback is either negative or positive. Here, the participant makes that adjudication. While this self-signification process was clearly preferred, its use was modified and narratives were assigned to positions on the map for two reasons – first, not all interviewees participated in the scenario planning (and so the conceptualizations may not have been their own), and second, time was often short (interviews were anywhere from 20-90 minutes in length, with the longest lengths reserved for larger group sizes).

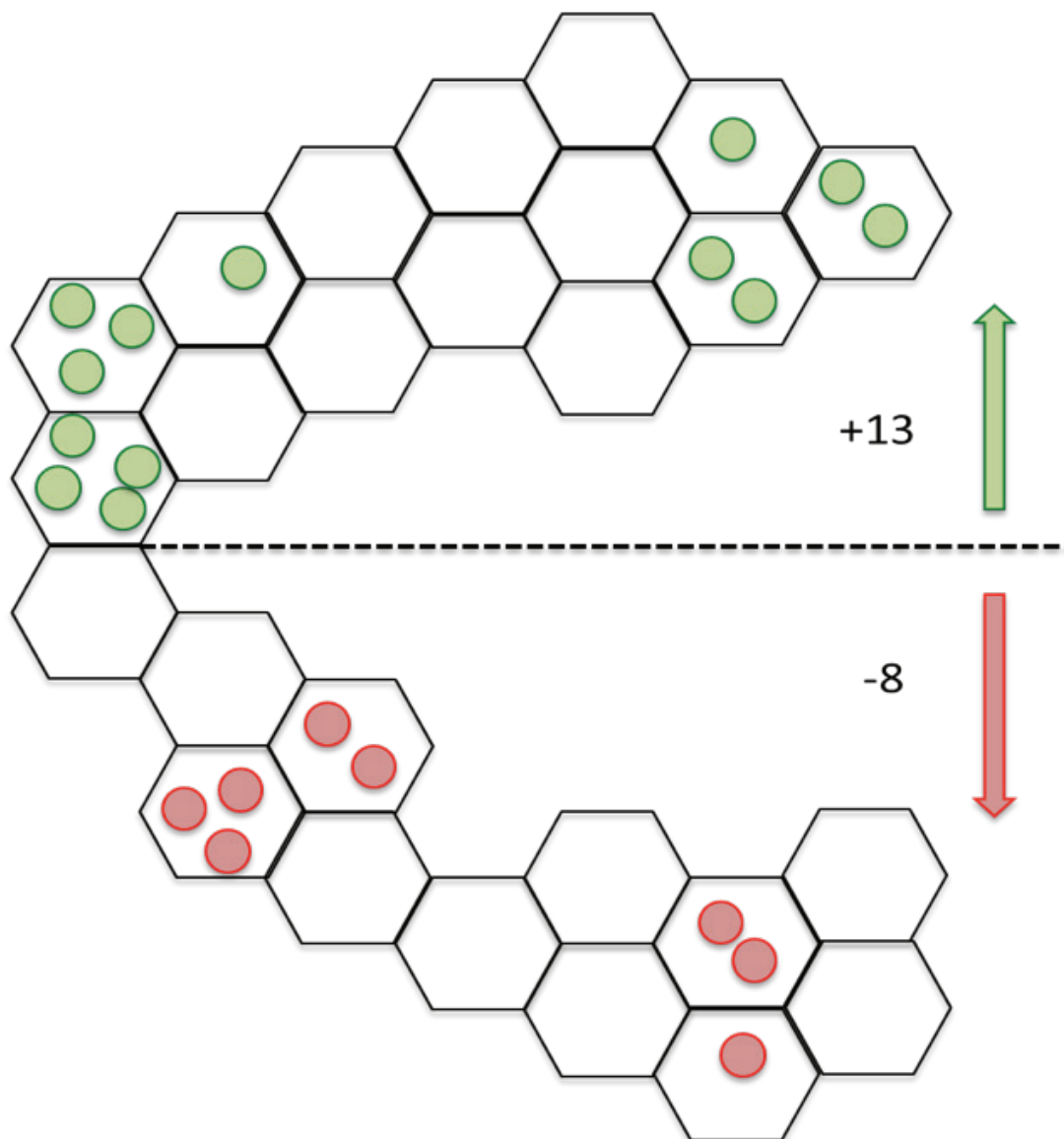
In Figure 67, dots signify where a given stakeholder thought his or her story matched one of the stories from the scenario planning process. As is indicated in the Figure, there were thirteen instances of narrative that reinforced a “utopian” path, and only eight that were indicative of a “dystopian” path. Moreover, it should be noted that five of the eight dystopian narratives came from just one site.

Figure 67. Results of self-signification process with fragmented micro-narratives, Spring 2012



For added clarity, the focus is on just the narratives that were dystopian and utopian, and abstract to just the simple graphic offered in Figure 68.

Figure 68. Final graph of micro-narrative process for determining the sense of the stakeholder community for whether improvement was occurring



These stories, and their locations, and the overall word clouds, were used to make some judgments about the perception of various stakeholder groups. To protect anonymity, this document does not report where certain stakeholder stories mapped in the scenario planning.

Each stakeholder group was asked to reflect on the contributions of other groups to the improvement process. In other words, parents, as an example, would be asked how they felt about the actions of 1)

students, 2) teachers, 3) administrators, 4) other community members, 5) the school board, and 6) the Montana Office of Public Instruction. While it is not particularly useful to report back on the “sentiment” analysis done of each group in each site (though many of those reflections have been shared privately with schools themselves in a briefing process that took place in March of 2013), it was surprising to the research team that the Montana OPI was, without exception at the group level (there were individual dissenters), perceived very favorably.

Each interview with various stakeholder groups was guided by a semi-structured interview outline.

Frazer word clouds are composed of multi-colored words on a white background; Lane Deer word clouds are composed of multi-colored words on a black background; and Pryor word clouds are composed of black words on a white background when this report is visualized digitally or in color print.

PERCEPTION IMPROVEMENT (STUDENTS)

Figure 69. Overall student perceptions for the Montana Schools of Promise

Overall	Frazer
	Lame Deer
	Pryor

Overall, the students in each school site perceived significant improvement. There was some concern in Lame Deer that there were still areas to improve, and that in some instances, the schools were not performing well, but students nevertheless claimed that they were seeing strong signs of growth. In general, students were the most positive stakeholder group interviewed; it may be that adults are shielding students from conversations that might indicate a messier implementation of reform efforts.

Overall word clouds are produced for each site, and some pulled quotes, without attribution to site, follow.

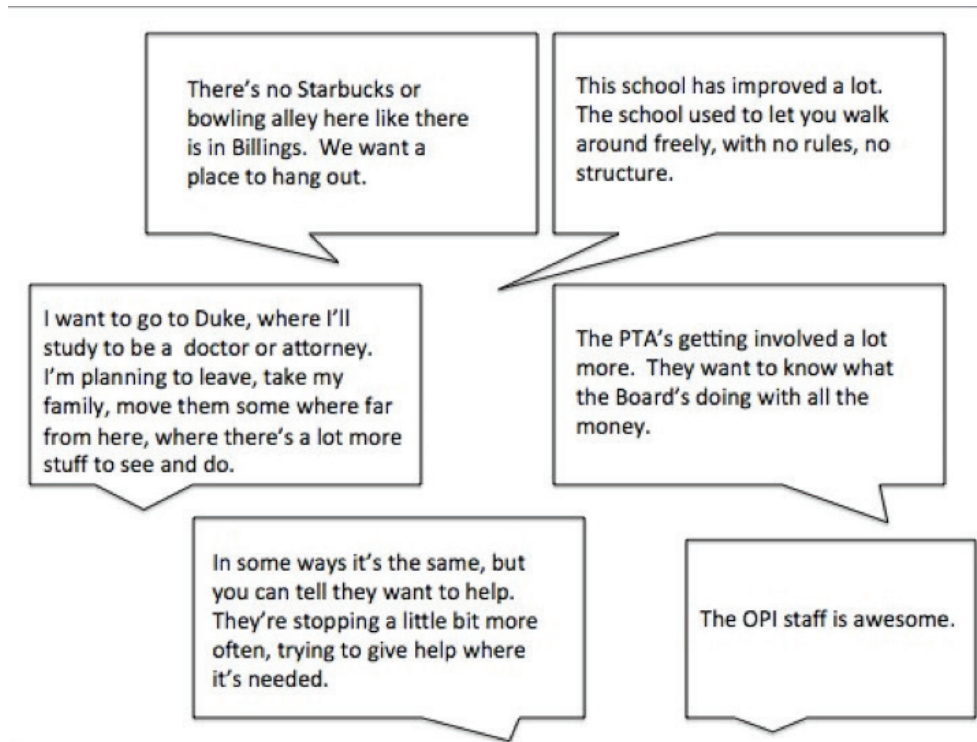
Figure 70. Student perception word cloud graph for Frazer, Spring 2012



SAMPLE OF COMMENTS FROM STUDENTS

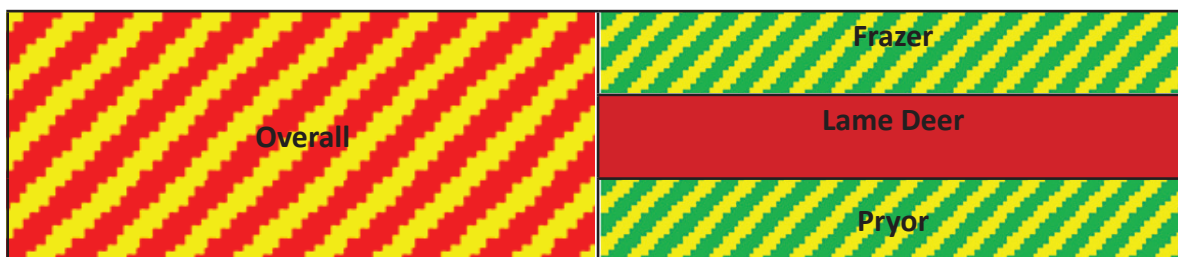
Representative comments – both positive and negative – are sampled in figure 73 below, from our student interviews.

Figure 73. Representative student comments, Spring 2012



PERCEPTION IMPROVEMENT (TEACHERS)

Figure 74. Overall teacher perceptions for the Montana Schools of Promise



Teacher perceptions were among the most critical of all stakeholder groups. They also tended to be the group that was most inward-facing, likely to attribute difficulties in schooling to themselves rather than other stakeholder groups (though they had criticisms of the contributions made by other stakeholders as well). Lane Deer teachers, in particular, expressed frustration with the lack of progress being seen in the schools. Some of their concerns are likely evident to readers upon a review of the word cloud generated by Lane Deer.

Figure 75. Teacher perception word cloud graph for Frazer, Spring 2012



Figure 76. Teacher perception word cloud graph for Lane Deer, Spring 2012

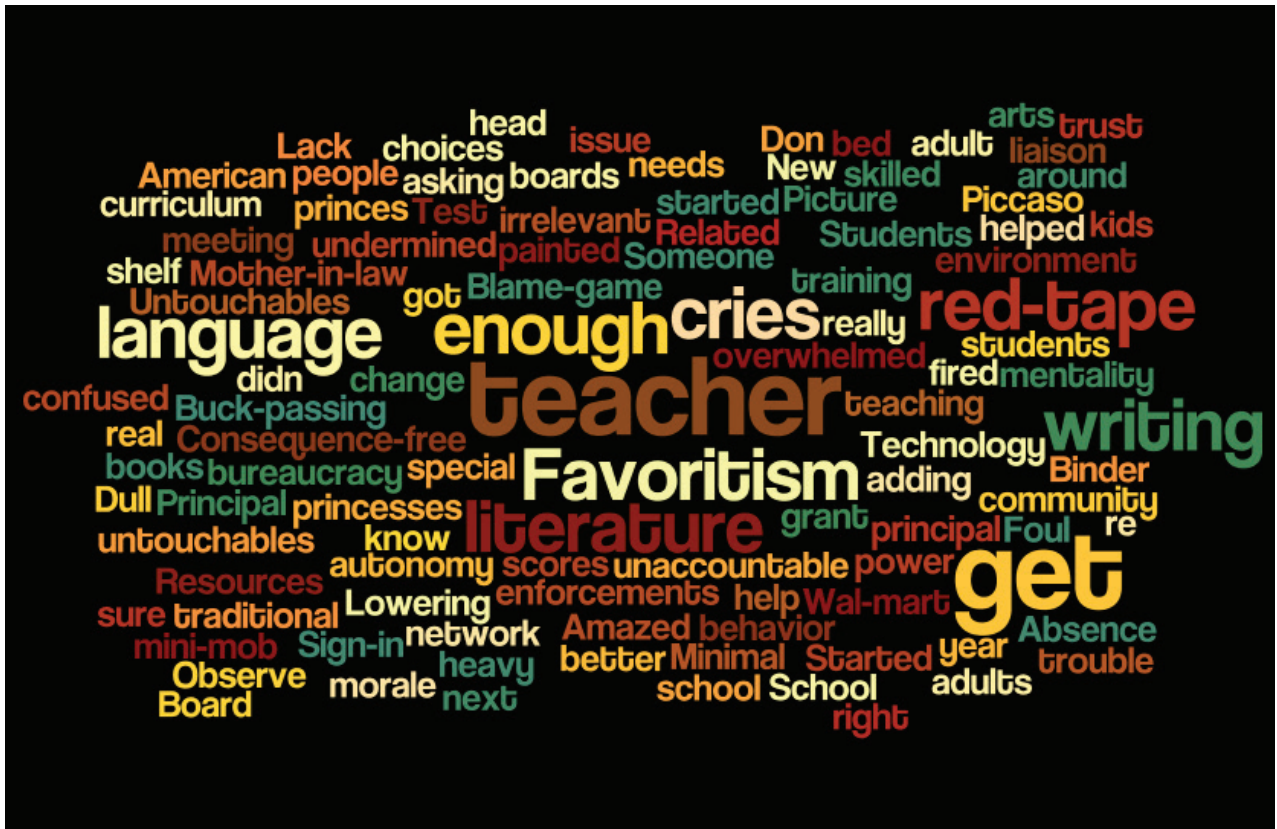


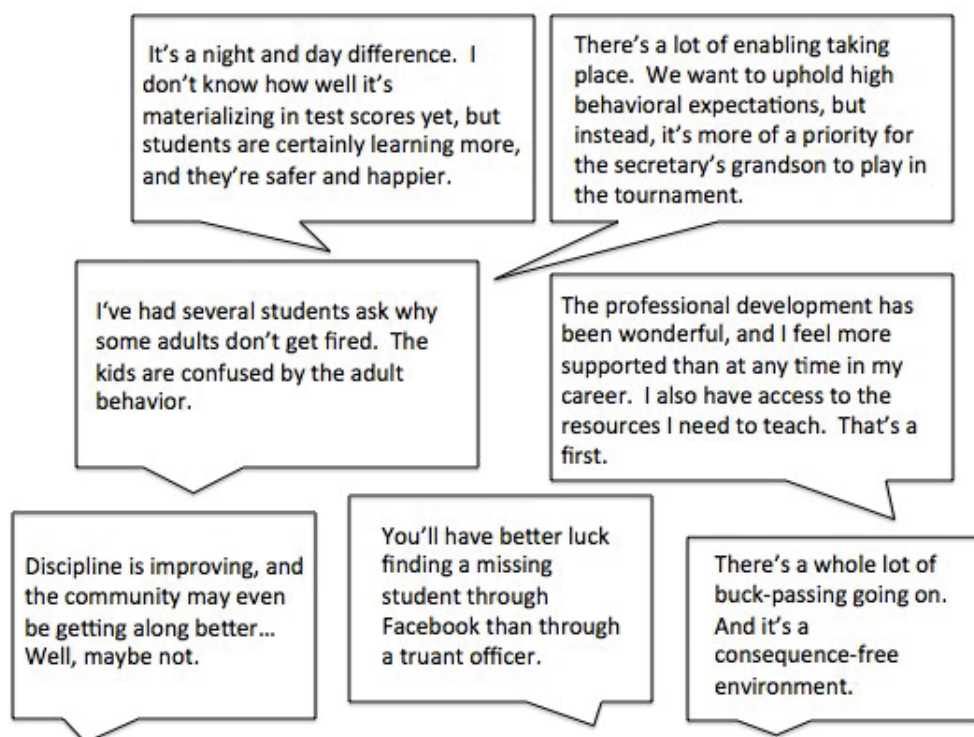
Figure 77. Teacher perception word cloud graph for Pryor, Spring 2012



SAMPLE OF COMMENTS FROM TEACHERS

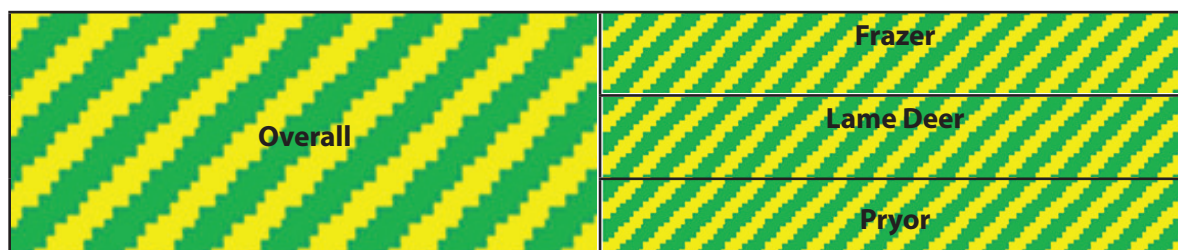
Representative comments – both positive and negative – are sampled in figure 78 below, from our teacher interviews.

Figure 78. Representative teacher comments, Spring 2012



PERCEPTION IMPROVEMENT (ADMINISTRATORS AND SUPPORT STAFF)

Figure 79. Overall administrator and support staff perceptions for the Montana Schools of Promise



Administrators, overall, were a positive group, indicating that they felt that the general trajectory was toward modest improvement. In both Frazer and Pryor, the administrators were more subdued in their statements about improvement, while Lane Deer administrators were more certain of improvement. The Lane Deer administrative view, however, was not shared by either teachers or community members.

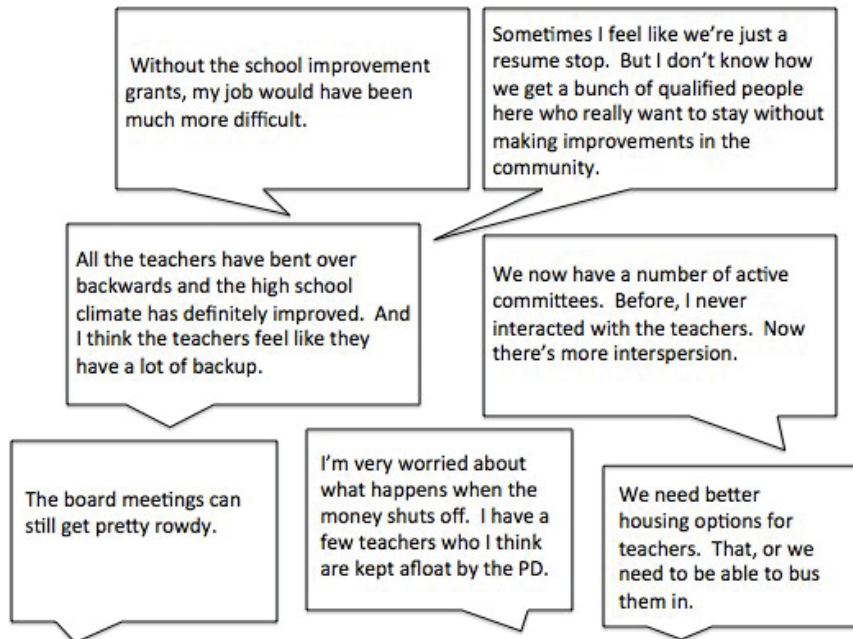
Figure 80. Administrator and support staff perception word cloud graph for Frazer, Spring 2012



SAMPLE OF COMMENTS FROM ADMINISTRATORS

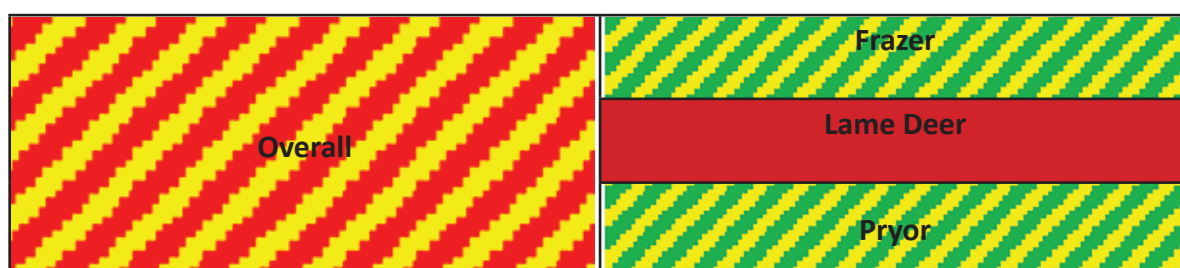
Representative comments – both positive and negative – are sampled in figure 83 below, from our administrator interviews.

Figure 83. Representative administrator comments, Spring 2012



PERCEPTION IMPROVEMENT (COMMUNITY AND PARENTS)

Figure 84. Overall community and parent perceptions for the Montana Schools of Promise



Community members and parents of school-age students were invited to participate in stakeholder interview groups. Frazer and Pryor community members and parents were generally appreciative of the reform efforts being made, with some criticisms, while Lame Deer community members expressed more significant dissatisfaction and frustration with improvement efforts, suggesting that the efforts were insufficient to improve the schools at an acceptable pace.

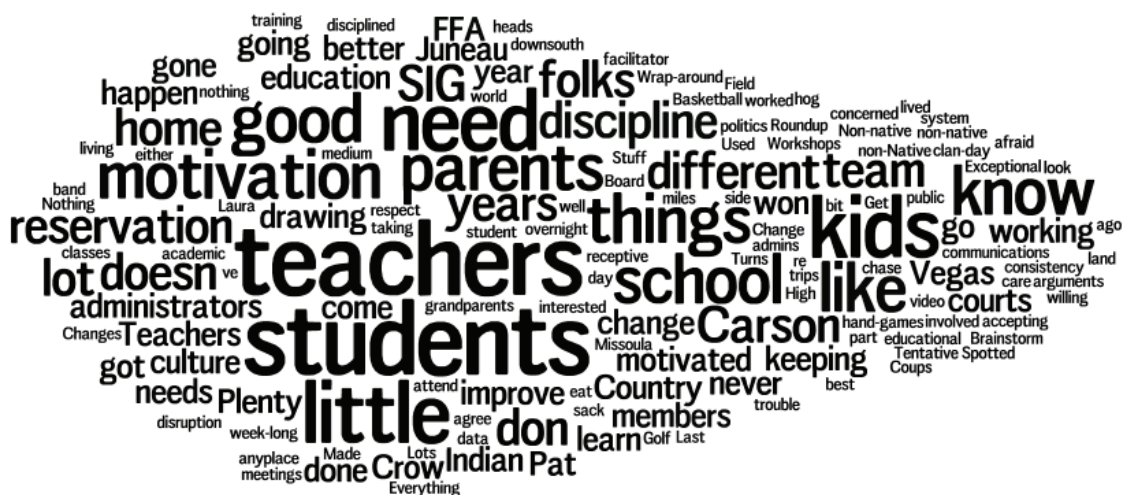
Figure 85. Community and parent perception word cloud graph for Frazer, Spring 2012



Figure 86. Community and parent perception word cloud graph for Lame Deer, Spring 2012



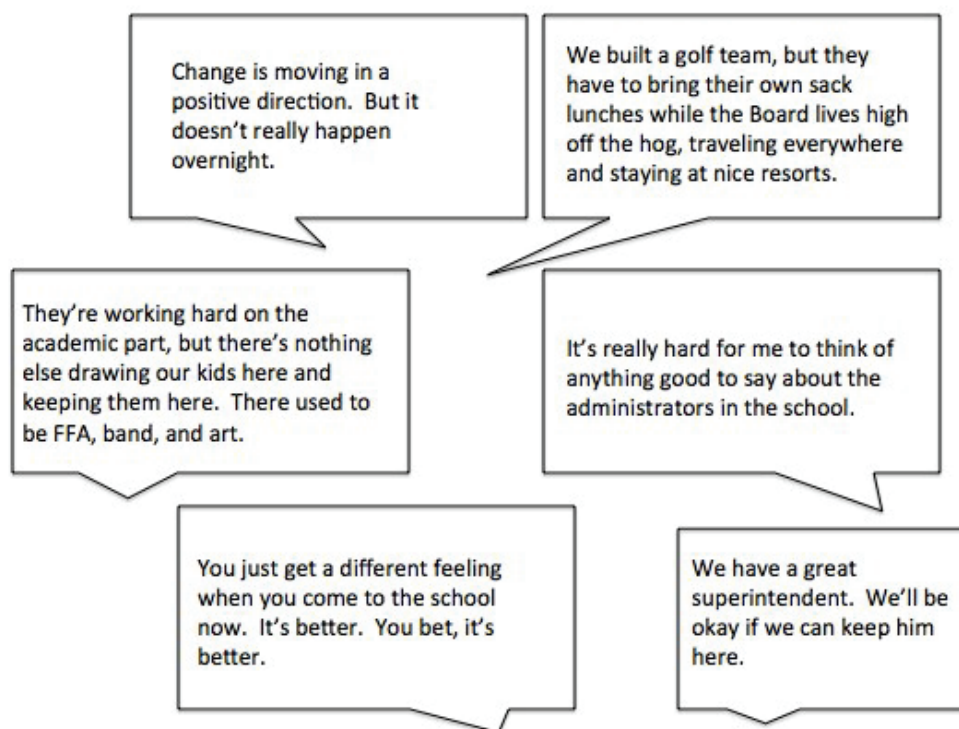
Figure 87. Community and parent perception word cloud graph for Pryor, Spring 2012



SAMPLE OF COMMENTS FROM COMMUNITY AND PARENTS

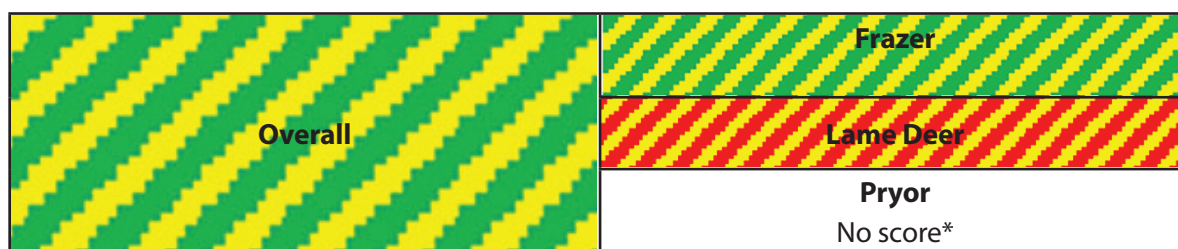
Representative comments – both positive and negative – are sampled in figure 88 below, from our community and parent interviews.

Figure 88. Representative student comments, Spring 2012



PERCEPTION IMPROVEMENT (SCHOOL BOARD MEMBERS)

Figure 89. Overall school board member perceptions for the Montana Schools of Promise



Frazer and Lame Deer school board members were split on their appreciation for the changes made in the SOP process, with Frazer generally enthusiastic, and Lame Deer more skeptical about whether the process yielded improvement. Unfortunately, no Pryor school board members were available during our site visits to conduct interviews. As a result, a rating for Pryor school board perception of the SOP process and its outcomes was not generated.

Figure 90. School board member perception word cloud graph for Frazer, Spring 2012



[illegible]

Representative comments – both positive and negative – are sampled in figure 92 below, from our student interviews.

Shop, carpentry, and mechanics are really important. The state test scores are up, but I don't think the ASVAB scores will be as good.

We need a lot more in the way of health and prevention. There's a wrap-around facilitator, but we probably need more of them, and more comprehensive services.

We're getting a lot of competition from the oil fields, and we're still keeping our enrollments up. So I think that's a good sign.

It's gotten a lot better. Our test scores are up and our discipline problems are down.

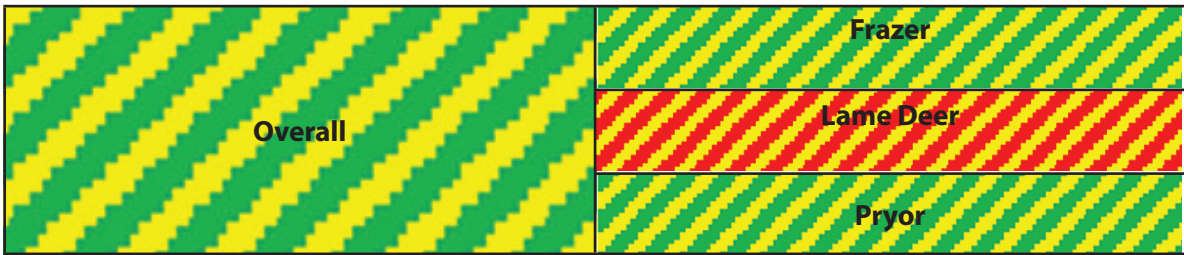
We're not acting in an integrated way. Teachers are separate from administrators. Parents are unhappy. We don't work together.

The school has to understand why it's important to quilt for an uncle's Memorial feast.

Pay attention to the moon and the seasons. One aide quit. Why? It was a season of madness.

PERCEPTION IMPROVEMENT (STATE)

Figure 89. Overall State (OPI) perceptions for the Montana Schools of Promise



As part of the SOP process, the Montana OPI placed several support team members at each school site for intensive coaching and project support. Present OPI team members were interviewed for their perceptions, included below.

Figure 90. State (OPI) perception word cloud graph for Frazer, Spring 2012



Figure 91. State (OPI) perception word cloud graph for Lane Deer, Spring 2012

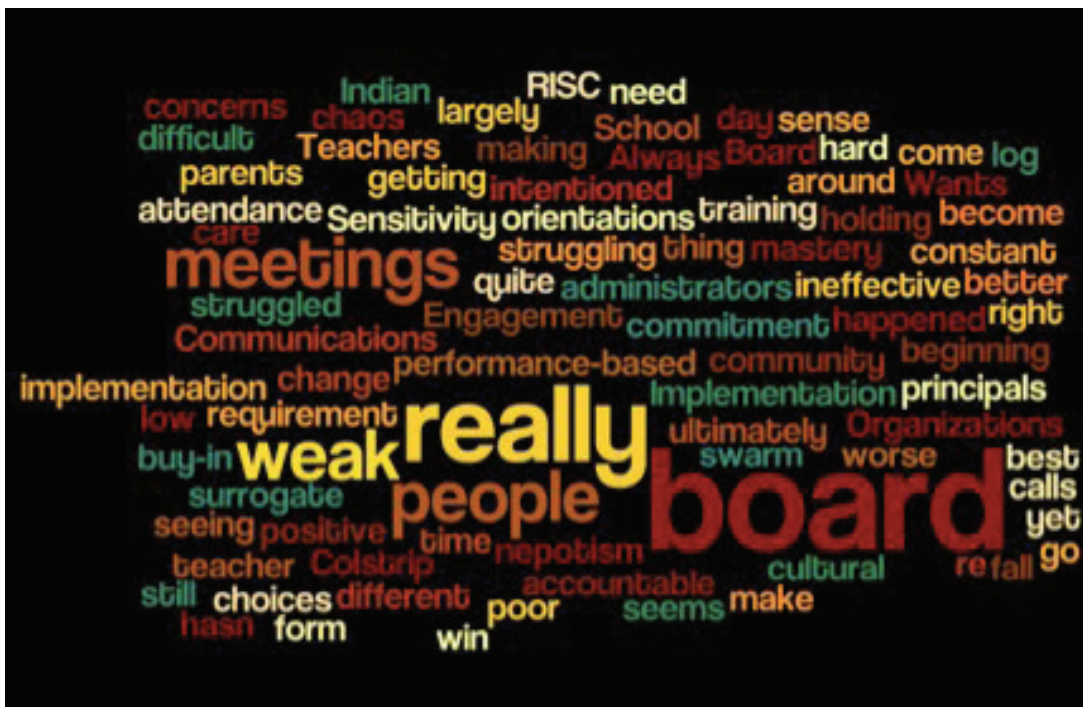


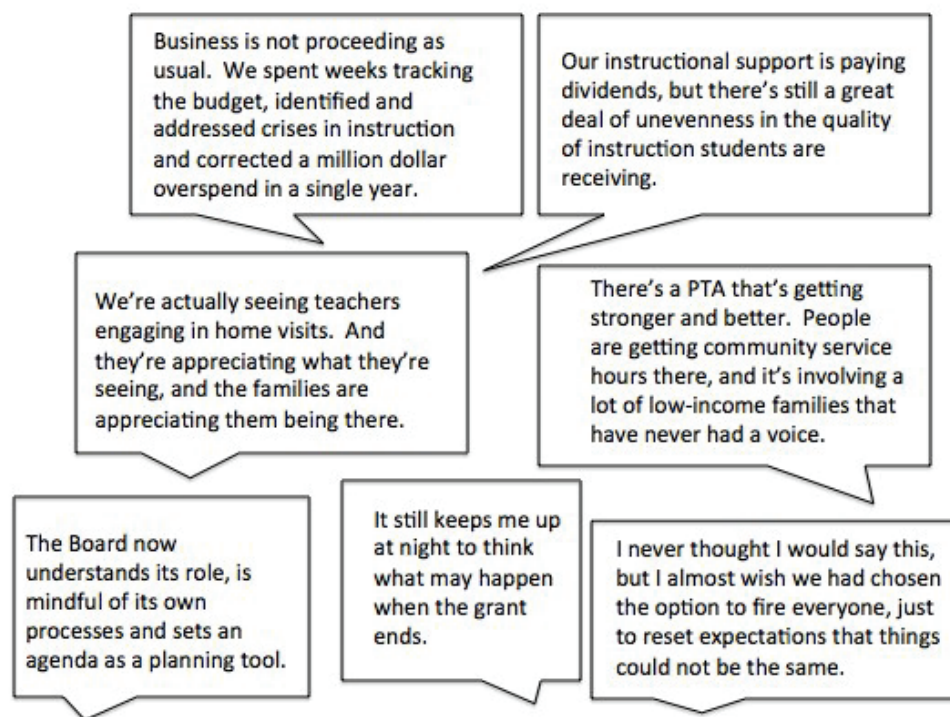
Figure 92. State (OPI) perception word cloud graph for Pryor, Spring 2012



SAMPLE STATE (OPI) COMMENTS

Representative comments – both positive and negative – are sampled in figure 93 below, from our state-level (OPI) interviews.

Figure 93. Representative state (OPI) comments, Spring 2012



Additional Results, Conclusions and Further Considerations

In addition to the information already described herein, multiple sources of other data were reviewed and examined in the evaluation process. For instance, results were examined from the reviews of OPI school support teams; evidence from researchers who have reviewed the SOP interventions; and popular media articles, often detailing the impressions of others who have visited the school sites extensively.

Though making casual statements about the totality of that evidence are difficult and probably ill advised, it is worth pointing out that the additional evidence roughly coincides with the appreciations of the present report. In other words, most evidence suggests general improvement, and more rapid progress in Frazer and Pryor schools than in Lame Deer.

Including the impressions of the school support teams in the overall rubric for performance would have been preferred. One of the principles for inclusion in the rubric, however, was that the measure had to be easy and of low-cost to replicate. This principle was important to honor because the role of the

rubric was that it be useful as an ongoing self-evaluation tool for the SOP. Unfortunately, the future of reviews by the school support teams, because of their relative expense and administrative intensity, did not meet the criteria for inclusion.

Additionally, relevant data was not able to be obtained on the performance of a quasi-experimental control group for each of the rubric's criteria. In the future, data collections should make comparisons on a range of factors to statistically or demographically alike schools if possible.

The role of student voice in the evaluation process was also considered. In addition to semi-structured interviews with groups of students themselves, data from the MyVoice Student Aspirations survey and other survey instruments from the Quaglia Institute was also examined as a result of support from the Pearson Foundation. Again, the results of that review roughly coincide with the overall impressions of the report. The survey data was excluded from the rubric, however, not because of expense or administrative intensity, as was the case with data from the school support teams, but because the surveys had been administered at such variant times within each of the schools. Going forward, the recommendation would be the inclusion of key survey data in the performance rubric, as well as the articulation of so-called "business rules" that would articulate the conditions, timing, and periodicity of surveying the students and other stakeholders called for by the instruments.

Looking at student well-being as reflected in the Youth Risk Behavior Surveys administered by the Montana OPI is also suggested. Presently, school level reports are not public, but aggregate view of student well-being in the SOP may be a worthy disaggregation of the data.

There were several interesting findings that did not naturally emerge in the context of describing a score or rating in the rubric and some of the most salient findings are offered here:

- With all stakeholder groups, a modified bias analysis was conducted. In other words, each group was asked to reflect on the strengths and weaknesses of each other stakeholder group (for instance, parents were asked what they thought of how students were reacting to and supporting the changes; how teachers were reacting to and supporting the changes; how administrators were reacting to and supporting the changes; how school board members were reacting to and supporting the changes; and how members of the state support staff were reacting to and supporting the changes). The bias analysis identified areas of perceived strengths and weaknesses. Those stakeholder groups that were perceived as weak by all other stakeholder groups had that information shared with them in private discussion. But perhaps more interestingly, the state support teams were recognized positively by all stakeholder groups – a surprise finding.
- Stakeholders, when reflecting on the effectiveness of the four models mandated by USED, universally cited the limitations of the models for their lack of a substantive and guiding research base. But there was also some curiosity about the model that would have required the dismissal of the staff. In all instances, it was not that they wished to see the actual removal of all or even many of the staff; they simply wanted to use the tool to, when re-hiring the staff back

for their assigned roles, set new expectations and re-configure some conditions of employment. There was a feeling that a few recalcitrant staff members were sub-optimizing performance. The sentiment was expressed not just about the faculty, but also about administrative staff as well.

- In each site (though to varying degrees) there are serious concerns about nepotism. In such small communities, not only is everyone likely to know one another, many times they are likely to be related. These facts make effective management somewhat more difficult, as already complicated relationships are stressed by the demands of familial obligations, too. Difficult conversations about the role of subjective, emotionally driven decisions in limiting the effectiveness of school-based efforts should be facilitated for the communities. Effective safeguards should be designed to ensure more objective decision-making.
- One of the most important features of the SOP improvement design was a focus on dealing with trauma in students' lives, and with trauma in the community more generally. Based on substantial evidence, the theory of action was that by helping to address the underlying emotional trauma experienced by communities residing in instances of concentrated, isolated, deep, generational poverty, with few historically positive interactions with a school system, dispositions toward school could be ameliorated, and academic achievement improved. Despite the relatively anemic uptake of the trauma informed training, there is widespread enthusiasm that it will become more popular and begin to yield more impact. As a result, it looks as though there will be community-based commitment to continuing and expanding the services going forward.
- In several instances, some of the data that appeared to point toward a negative trend were questioned. For instance, in one of the school sites that experienced a decline in graduation rates over the period of the grant, a few observers hazarded that perhaps the reason graduation rates appeared to be going down was merely because the historical data were inaccurate. There was some concern expressed that historic graduation rates were artificially high, masking the true numbers of students dropping out. Some went so far as to say "we're probably graduating a higher percentage of students than ever before; it's just that now, we're being honest about the rate."
- Cultural ways of knowing and understanding may also contribute to the performance dynamics of the schools and communities in the SOP.

Some of these features are interesting side notes, and some are avenues for further inquiry, potentially through research.

One area of research might be the reasons for the variant rates of improvement among the three schools. Lane Deer, struggling to boast clear signs of improvement, for instance, often cites the role of school choice in suppressing academic performance in the school. More motivated students, according to the hypothesis, elect to go to school in Colstrip (where academic achievement for American Indian students on the MontCAS is indeed higher, though other outcomes were not exhaustively examined

for American Indian students in Colstrip schools), or at the private St. Labre school. Clearer pictures of how schools in at-risk communities, and affected by negative selection pressures, manage to raise outcomes should be generated for the faculty, staff, and communities attempting to navigate to such improvement, though unfortunately, precious few examples of success against such pressures exist. Further research should be done to appreciate the potential contribution of school choice to the less robust outcomes in Lame Deer, and to suggest more creative, nuanced responses.

To return to the original question of whether the moneys spent helped 1) to ameliorate the educational outcomes for American Indian students, and 2), to produce new jobs that contribute as much to the economy as the investment in those jobs while abetting the first goal, only casual responses to each can be offered. With regard to the first, the present evaluation would suggest that resources were well spent and contributed to educational improvements, though again readers are reminded that the present evaluation was not a causal research study, and the suggestion is that one be undertaken in order to drive more definite statements about the grant's impact. As for the second, it was clear that a number of jobs were created as a result of the investment, but it is unclear what the net return on the investment actually was for the United States', or Montana's, economy. Any ultimate analysis, though, should also account for the number of contributing workers that may have been created as a result of improved schooling opportunity. If more graduates of the schools are produced, and more contribute meaningfully to an economy, there may be downstream investment impacts that need to be considered in any robust return on investment analysis.

In all, investments and efforts in the Montana Schools of Promise appear to have benefitted the students and communities they were designed to assist. Though much diligence and continued effort will be necessary to sustain and accelerate the observable progress, stakeholders appear poised to answer the challenge to address the needs of young people in their communities by provisioning more robust and meaningful educational opportunities.

APPENDIX A: RUBRIC (PART I. ASSESSMENT OUTCOMES)

	Red	Amber/Red	Amber/Green	Green
Reading improvement	The percentage of students who are scoring proficient or higher has declined. The percentage of students performing at the novice level has grown.	Overall, there may be little to no improvement in the percentage of students scoring proficient or higher. There may, however, be a decline in the percentage of students scoring at the novice level.	More students are scoring proficient, and fewer students are scoring at the novice level. Improvement, however, may not outpace comparison schools.	More students are scoring proficient, and fewer students are scoring at the novice level, and the improvement exceeds the rate of improvement in comparison schools.
	Red	Amber/Red	Amber/Green	Green
Mathematics improvement	The percentage of students who are scoring proficient or higher has declined. The percentage of students performing at the novice level has grown.	Overall, there may be little to no improvement in the percentage of students scoring proficient or higher. There may, however, be a decline in the percentage of students scoring at the novice level.	More students are scoring proficient, and fewer students are scoring at the novice level. Improvement, however, may not outpace comparison schools.	More students are scoring proficient, and fewer students are scoring at the novice level, and the improvement exceeds the rate of improvement in comparison schools.
	Red	Amber/Red	Amber/Green	Green
Science improvement	The percentage of students who are scoring proficient or higher has declined. The percentage of students performing at the novice level has grown.	Overall, there may be little to no improvement in the percentage of students scoring proficient or higher. There may, however, be a decline in the percentage of students scoring at the novice level.	More students are scoring proficient, and fewer students are scoring at the novice level. Improvement, however, may not outpace comparison schools.	More students are scoring proficient, and fewer students are scoring at the novice level, and the improvement exceeds the rate of improvement in comparison schools.

Appendix A: Rubric (Part II. Other Academic Indicators)

	Red	Amber/Red	Amber/Green	Green
Graduation rate improvement	The percentage of students who are graduating has declined.	There is no improvement in the percentage of students graduating, or irregular and erratic patterns of improvement over the time interval considered.	There is modest improvement in the percentage of students graduating, or generalized trend improvement over the time interval examined.	Higher percentages of students are graduating from high school.
Attendance improvement	Demonstrably fewer students are attending school, and the rate of attendance decline is steeper than in comparison groups.	There is no improvement in the percentage of students graduating, or irregular and erratic patterns of improvement over the time interval considered.	Attendance trends have improved, but as not as sharply as in comparison schools (eg state and/or historically and demographically similar schools)	Attendance trends have improved, and at rates that exceed the attendance improvement in comparison schools.
Suspension improvement	More students are being suspended, and at a rate that exceeds comparison schools.	Suspension rates are increasing, but not at a faster rate than comparison groups.	Suspension rates are declining, but not at a rate that outpaces comparison groups.	Suspension rates are declining at a rate that exceeds the rate of decline in comparison schools.
Expulsion improvement	More students are being expelled, and the rate is increasing faster than comparison groups.	Expulsion rates have not improved, and/or are erratic over the time interval. Expulsion rates may be higher, but may not be increasing at a rate that exceeds that of comparison schools.	Expulsion rates are declining, generally, but not in a way that clearly outpaces comparison schools.	Expulsion rates are declining at a rate that exceeds the rate of decline in comparison schools.
Student engagement improvement	Scores of student engagement, as measured and visualized, are declining.	Scores of student engagement are not improving, or are declining, but not as rapidly as comparison schools.	Student engagement rates are increasing, but at a rate that is slower than comparison groups.	Student engagement rates are increasing at a rate that exceeds the increases seen in comparison groups.

Appendix A: Rubric (Part III. Perception data and findings)

Perception improvement (students)	The stakeholder group perceives that the school is in decline, performing worse than it was historically.	The stakeholder group perceives little improvement in the school, and is not optimistic about the prospects for future success.	The stakeholder group may or may not perceive improvement in the school, but is optimistic about the prospects for future success.	The stakeholder group perceives improvement in the school, and is optimistic about the prospects for future success.
	Red	Amber/Red	Amber/Green	Green
Perception improvement (teachers)	The stakeholder group perceives that the school is in decline, performing worse than it was historically.	The stakeholder group perceives little improvement in the school, and is not optimistic about the prospects for future success.	The stakeholder group may or may not perceive improvement in the school, but is optimistic about the prospects for future success.	The stakeholder group perceives improvement in the school, and is optimistic about the prospects for future success.
	Red	Amber/Red	Amber/Green	Green
Perception improvement (administrators and support staff)	The stakeholder group perceives that the school is in decline, performing worse than it was historically.	The stakeholder group perceives little improvement in the school, and is not optimistic about the prospects for future success.	The stakeholder group may or may not perceive improvement in the school, but is optimistic about the prospects for future success.	The stakeholder group perceives improvement in the school, and is optimistic about the prospects for future success.
	Red	Amber/Red	Amber/Green	Green
Perception improvement (community)	The stakeholder group perceives that the school is in decline, performing worse than it was historically.	The stakeholder group perceives little improvement in the school, and is not optimistic about the prospects for future success.	The stakeholder group may or may not perceive improvement in the school, but is optimistic about the prospects for future success.	The stakeholder group perceives improvement in the school, and is optimistic about the prospects for future success.
	Red	Amber/Red	Amber/Green	Green
Perception improvement (board members)	The stakeholder group perceives that the school is in decline, performing worse than it was historically.	The stakeholder group perceives little improvement in the school, and is not optimistic about the prospects for future success.	The stakeholder group may or may not perceive improvement in the school, but is optimistic about the prospects for future success.	The stakeholder group perceives improvement in the school, and is optimistic about the prospects for future success.
	Red	Amber/Red	Amber/Green	Green
Perception improvement (state)	The stakeholder group perceives that the school is in decline, performing worse than it was historically.	The stakeholder group perceives little improvement in the school, and is not optimistic about the prospects for future success.	The stakeholder group may or may not perceive improvement in the school, but is optimistic about the prospects for future success.	The stakeholder group perceives improvement in the school, and is optimistic about the prospects for future success.



Montana
Schools of
Promise
opi.mt.gov

The Office of Public Instruction is committed to equal employment opportunity and nondiscriminatory access to all our programs and services, and will take necessary and appropriate steps to insure that the workplace and OPI programs and services are free of discrimination and sexual harassment. For information or to file a complaint, contact OPI Title IX/EEO Coordinator at (406) 444-3161 or opipersonnel@mt.gov.

Copies of this public document were published at an estimated cost of \$. per copy, for a total cost of \$.00, which includes \$.00 for printing and \$0.00 for distribution